



PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 39237  
LENGTH: 31  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC006538.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5  
US-09-864-761-39237  
Query Match 34.3%; Score 48; DB 10; Length 31;  
Best Local Similarity 50.0%; Pred. No. 0.69;  
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
OY 8 ENVPGHERMGRGRTSSKELA 27  
DB 4 EPVPGHARPGHRRSGSAA 23  
RESULT 2  
US-09-815-242-11988  
Sequence 11988, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlson, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE OF INVENTION: Prokaryotes  
FILE REFERENCE: EITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11988  
LENGTH: 206  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa

US-09-815-242-11988  
Query Match 31.1%; Score 43.5; DB 10; Length 206;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 11; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
OY 2 LEAKHRENVPGHERMGR 22  
DB 28 LDSKCAENVPGHGRGRRLS 49  
RESULT 3  
US-09-925-300-1493  
Sequence 1493, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben,  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1493  
LENGTH: 88  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (10)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (29)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (73)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1493  
Query Match 30.7%; Score 43; DB 10; Length 88;  
Best Local Similarity 47.1%; Pred. No. 11;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
OY 2 LEAKHRENVPGHERMGR 18  
DB 46 IEKSWRELVPCHKMSQ 62  
RESULT 4  
US-09-864-761-35901  
Sequence 35901, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
FILE REFERENCE: Aecmca-x-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 35901  
LENGTH: 102  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL096678.8  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EST\_HUMAN HIT: BE792924.1, EVALUATE 1.40e-01  
S-09-864-761-35901  
Query Match 30.7%; Score 43; DB 10; Length 102;  
Best Local Similarity 37.5%; Pred. No. 13;  
Matches 9; Conservative 3; Mismatches 12; Indels 0; Gaps 0;  
QY 2 LEAKHENVGHERMGRGRTSSKE 25  
DB 4 LEIKKEPEAGHAEERGETETDQ 27  
RESULT 5  
US-09-811-284-208  
Sequence 208, Application US/09811284  
Patent No. US20020058306A1  
GENERAL INFORMATION:  
APPLICANT: Vogeli, Gabriel  
TITLE OF INVENTION: No. US20020058306A1 G Protein-Coupled Receptors  
FILE REFERENCE: 00167US1  
CURRENT FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: US/09/811,284  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/189,783  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/189,907  
PRIOR FILING DATE: 2000-03-16

PRIOR APPLICATION NUMBER: 60/189,918  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/189,960  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/189,917  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/192,945  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,916  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,923  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,933  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,830  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,234  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: 60/192,155  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: 60/192,935  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 258  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 208  
LENGTH: 165  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-811-284-208  
Query Match 30.7%; Score 43; DB 10; Length 165;  
Best Local Similarity 42.1%; Pred. No. 21;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;  
QY 5 KRENVGHERMGRGRTSS 23  
DB 46 KRESLPQNRSSSGTSYS 64  
RESULT 6  
US-09-864-761-34158  
Sequence 34158, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica X-1  
CURRENT FILING DATE: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: US 60/24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34158
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000089.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HB100, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.5
; OTHER INFORMATION: EST HUMAN HIT: AU131304.1, EVALUATE 2.00e-19
; OTHER INFORMATION: SWISSPROT HIT: Q95241, EVALUATE 2.00e-20
; US-09-864-761-34158

Query Match
Best Local Similarity 30.0%; Score 42; DB 10; Length 44;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLEAKHRENV 10
Db 33 RLEAKHRENV 42

RESULT 7
; US-09-815-242-11860
; Sequence 11860, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11860
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-11860

Query Match
Best Local Similarity 30.0%; Score 42; DB 10; Length 286;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 RLEAKHRENVGHE 14
Db 8 LPAQHODQRPGE 20

RESULT 8
; US-09-815-242-5083
; Sequence 5083, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5083
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-5083

Query Match
Best Local Similarity 44.4%; Score 42; DB 10; Length 327;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 RLEAKHRENVGHEMRGR 18
Db 48 RADPRHREALCEMRRLMR 65

RESULT 9
```

```
US-09-864-761-37938
; Sequence 37938, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37938
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005074.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EST_HUMAN HIT: AW976211.1, EVALUATE 2.00e-91
; OTHER INFORMATION: SWISSPROT HIT: P53935, EVALUATE 2.00e-03
US-09-864-761-37938

Query Match 30.0%; Score 42; DB 10; Length 560;
Best Local Similarity 41.7%; Pred. No. 1.1e+02;
Matches 10; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

QY 2 LEAKHRENVGHERMGRTSSKE 25
DB 497 MEAKNKN--GKVENGLGKTDRKK 518

RESULT 10
US-09-864-761-34546
; Sequence 34546, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34546
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005089.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
```

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7  
OTHER INFORMATION: SWISSPROT HIT: P53935, EVALUATE 2.00e-03  
OTHER INFORMATION: EST\_HUMAN HIT: AM976211.1, EVALUATE 2.00e-91  
US-09-864-761-34546

Query Match 30.0%; Score 42; DB 10; Length 572;  
Best Local Similarity 41.7%; Pred. No. 1.1e+02;

Matches 10; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

OY 2 LEAKHRENVGHERMGRGRSSKE 25

Db 504 MEAKNKEV--GKVENGLGKTDKK 525

RESULT 11

US-09-810-264-38  
Sequence 38, Application US/09810264  
Patent No. US20020076775A1  
GENERAL INFORMATION:  
APPLICANT: Crane, Virginia C.  
APPLICANT: Famodu, Omolayo O.  
APPLICANT: Hu, Xu  
APPLICANT: Lu, Guohua  
APPLICANT: Zhang, Lingyu  
TITLE OF INVENTION: WRKY Transcription Factors and Methods  
TITLE OF INVENTION: Of use  
FILE REFERENCE: 1183  
CURRENT APPLICATION NUMBER: US/09/810,264  
CURRENT FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: US 60/190,467  
PRIOR FILING DATE: 2000-03-17  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: PASTESEQ for Windows Version 4.0  
SEQ ID NO 38  
LENGTH: 583  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-810-264-38

Query Match 30.0%; Score 42; DB 10; Length 583;  
Best Local Similarity 38.1%; Pred. No. 1.1e+02;

Matches 8; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

OY 3 EAKHRENVGHERMGRGRSS 23

Db 438 EKHNEHYPARSSGSGSSGS 458

RESULT 12

US-09-794-927-10  
Sequence 10, Application US/09794927  
Patent No. US20010016324A1  
GENERAL INFORMATION:  
APPLICANT: Gurney, Mark E.  
APPLICANT: Bienkowski, Michael J.  
APPLICANT: Heinrichson, Robert L.  
APPLICANT: Parodi, Luis A.  
APPLICANT: Yan, Riqiang  
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND  
TITLE OF INVENTION: USES  
FILE REFERENCE: 28341/6280FG  
CURRENT APPLICATION NUMBER: US/09/794,927  
CURRENT FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: 09/416,901  
PRIOR FILING DATE: 1999-10-13  
PRIOR APPLICATION NUMBER: 60/155,493  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 09/404,133  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: PCT/US99/20881  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 60/101,594

PRIOR FILING DATE: 1998-09-24  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 10  
LENGTH: 695  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-794-927-10

Query Match 30.0%; Score 42; DB 10; Length 695;  
Best Local Similarity 80.0%; Pred. No. 1.4e+02;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLEAKHRENV 10

Db 322 RLEAKHRERM 331

RESULT 13

US-09-794-927-12  
Sequence 12, Application US/09794927  
Patent No. US20010016324A1  
GENERAL INFORMATION:  
APPLICANT: Gurney, Mark E.  
APPLICANT: Bienkowski, Michael J.  
APPLICANT: Heinrichson, Robert L.  
APPLICANT: Parodi, Luis A.  
APPLICANT: Yan, Riqiang  
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND  
TITLE OF INVENTION: USES  
FILE REFERENCE: 28341/6280FG  
CURRENT APPLICATION NUMBER: US/09/794,927  
CURRENT FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: 09/416,901  
PRIOR FILING DATE: 1999-10-13  
PRIOR APPLICATION NUMBER: 60/155,493  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 09/404,133  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: PCT/US99/20881  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 60/101,594  
PRIOR FILING DATE: 1998-09-24  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 12  
LENGTH: 695  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-794-927-12

Query Match 30.0%; Score 42; DB 10; Length 695;  
Best Local Similarity 80.0%; Pred. No. 1.4e+02;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLEAKHRENV 10

Db 322 RLEAKHRERM 331

RESULT 14

US-09-794-927-14  
Sequence 14, Application US/09794927  
Patent No. US20010016324A1  
GENERAL INFORMATION:  
APPLICANT: Gurney, Mark E.  
APPLICANT: Bienkowski, Michael J.  
APPLICANT: Heinrichson, Robert L.  
APPLICANT: Parodi, Luis A.  
APPLICANT: Yan, Riqiang  
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND  
TITLE OF INVENTION: USES

Search completed: October 29, 2002, 14:07:51  
 Job time : 40 secs

```

; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 28341/6280RG
; CURRENT APPLICATION NUMBER: US/09/794,927
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 14
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-794-927-14

```

## Query Match

30.0%; Score 42; DB 10; Length 695;

Best Local Similarity 80.0%; Pred. No. 1.4e+02;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLEAKHRENV 10  
 ||||||| :

Db 322 RLEAKHRENV 331

## RESULT 15

```

; US-09-795-847-10
; Sequence 10, Application US/09795847
; Patent No. US20010018208A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrichson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yao, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES
; FILE REFERENCE: 28341/6280DE
; CURRENT APPLICATION NUMBER: US/09/795,847
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-795-847-10

```

## Query Match

30.0%; Score 42; DB 10; Length 695;

Best Local Similarity 80.0%; Pred. No. 1.4e+02;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLEAKHRENV 10  
 ||||||| :

Db 322 RLEAKHRENV 331





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 13:57.43 ; Search time 26 Seconds  
(without alignments) 30.555 Million cell updates/sec

Title: US-09-674-913a-2  
Perfect score: 140  
Sequence: 1 RLEAKHRENVGHERMGRGRTSKELA 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

tal number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	ID	Description
1	109	77.9	25	2	US-08-726-306A-61 Sequence 61, Appl
2	109	77.9	25	2	US-08-726-306A-87 Sequence 87, Appl
3	50	35.7	26	1	US-07-942-245-321 Sequence 321, App
4	47	33.6	10	2	US-08-726-306A-11 Sequence 1, Appl
5	46	32.9	26	1	US-07-942-245-312 Sequence 312, App
6	45.5	32.5	537	4	US-09-134-001C-4091 Sequence 4091, Ap
7	44	31.4	26	1	US-07-942-245-314 Sequence 314, App
8	44	31.4	26	1	US-07-942-245-315 Sequence 315, App
9	44	31.4	26	1	US-07-942-245-349 Sequence 349, App
10	44	31.4	295	4	US-09-026-482B-2 Sequence 2, Appl
11	44	31.4	617	4	US-09-134-001C-4012 Sequence 4012, Ap
12	43	30.7	26	1	US-07-942-245-311 Sequence 311, Appl
13	43	30.7	119	5	US-08-581-528A-7 Sequence 7, Appl
14	43	30.7	119	5	PCT-US94-07799-7 Sequence 6, Appl
15	43	30.7	161	2	US-08-581-528A-6 Sequence 6, Appl
16	43	30.7	161	5	PCT-US94-07799-6 Sequence 6, Appl
17	43	30.7	855	2	US-08-816-693A-2 Sequence 2, Appl
18	43	30.7	855	3	US-08-885-291-2 Sequence 2, Appl
19	43	30.7	855	4	US-09-496-672-2 Sequence 2, Appl
20	42	30.0	17	5	PCT-US92-09070-8 Sequence 8, Appl
21	42	30.0	26	1	US-07-942-245-302 Sequence 302, App
22	42	30.0	30	2	US-08-726-306A-59 Sequence 59, Appl
23	42	30.0	38	1	US-08-253-155A-45 Sequence 45, Appl
24	42	30.0	40	5	PCT-US92-09070-5 Sequence 5, Appl
25	42	30.0	50	4	US-09-314-268-159 Sequence 159, Appl
26	42	30.0	68	1	US-07-756-250-10 Sequence 10, Appl
27	42	30.0	129	1	US-08-360-914B-15 Sequence 15, Appl

28	42	30.0	129	1	US-08-741-589A-13 Sequence 13, Appl
29	42	30.0	155	1	US-08-339-152A-23 Sequence 23, Appl
30	42	30.0	155	2	US-08-007-999B-10 Sequence 10, Appl
31	42	30.0	155	2	US-08-689-276A-10 Sequence 10, Appl
32	42	30.0	160	2	US-08-726-306A-179 Sequence 179, App
33	42	30.0	276	1	US-07-882-329-2 Sequence 2, Appl
34	42	30.0	276	1	US-08-183-213-2 Sequence 2, Appl
35	42	30.0	294	1	US-08-362-670B-2 Sequence 2, Appl
36	42	30.0	294	3	US-08-333-576C-2 Sequence 2, Appl
37	42	30.0	294	4	US-08-808-324-2 Sequence 2, Appl
38	42	30.0	294	5	PCT-US94-14030A-2 Sequence 2, Appl
39	42	30.0	388	1	US-08-362-670B-34 Sequence 34, Appl
40	42	30.0	388	3	US-08-333-576C-34 Sequence 34, Appl
41	42	30.0	388	4	US-08-808-324-34 Sequence 34, Appl
42	42	30.0	388	5	PCT-US94-14030A-34 Sequence 34, Appl
43	42	30.0	411	1	US-08-362-670B-28 Sequence 28, Appl
44	42	30.0	411	3	US-08-333-576C-28 Sequence 28, Appl
45	42	30.0	411	4	US-08-808-324-28 Sequence 28, Appl

ALIGNMENTS

\* RESULT 1  
US-08-726-306A-61  
Sequence 61, Application US/08726306A  
Patent No. 5958684  
GENERAL INFORMATION:  
APPLICANT: van Leeuwen, Frederik Willem  
APPLICANT: Burbach, Johannes Peter Henri  
APPLICANT: Grosveld, Franklin G.  
TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS  
NUMBER OF SEQUENCES: 189  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1 Financial Center  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 1.44 MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/726,306A  
FILING DATE: 02-Oct-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 95/20080.4  
FILING DATE: 02-Oct-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/009,832  
FILING DATE: 01-Jan-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams, Ph D., Kathleen M.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 345-9100  
TELEFAX: (617) 345-9111  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-726-306A-61  
Query Match 77.9%; Score 109; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.3e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

?
? REFERENCE/DOCKET NUMBER: 96_048-A (3255/00784)
?
? TELECOMMUNICATION INFORMATION:
?
? TELEPHONE: (617) 345-9100
?
? TELEFAX: (617) 345-9111
?
? INFORMATION FOR SEQ ID NO: 1:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 10 amino acids
?
? TYPE: amino acid
?
? STRANDEDNESS: single
?
? TOPOLOGY: unknown
?
? MOLECULE TYPE: peptide
?
? OS-08-726-306A-1

```

Query Match	33.6%;	Score 47;	DB 2;	Length 10;
Best Local Similarity	100.0%;	Pred. No. 0.084;		
Matches 10;	Conservative 0;	Mismatches 0;	Indels	

QY	18	RGR	TSS	SKEL	27
25	1	RGR	TSS	SKEL	10

```

1      RESULT 5 -
2      US-07-942-245-312
3      : Sequence 312, Application US/07942245
4      : Patent No. 5639641
5      : GENERAL INFORMATION:
6      : APPLICANT: PEDERSEN, Jan T.
7      : APPLICANT: SEARLE, Stephen M.J.
8      : APPLICANT: REES, Anthony R.
9      : APPLICANT: ROGUSKA, Michael A.
10     : APPLICANT: GULD, Braydon C.
11     : TITLE OF INVENTION: SURFACE RESIDUE VENEERING OF RODENT
12     : TITLE OF INVENTION: ANTIBODIES
13     : NUMBER OF SEQUENCES: 522
14     : CORRESPONDENCE ADDRESS:
15     : ADDRESSEE: Sugrute, Mfon, Zinn, Macpeak & Seas
16     : STREET: 2100 Pennsylvania Avenue, N.W.
17     : City: Washington
18     : STATE: D.C.
19     : COUNTRY: United States
20     : ZIP: 20037-3202
21     : COMPUTER READABLE FORM:
22     : MEDIUM TYPE: Floppy disk
23     : COMPUTER: HP 9000/700 Workstation
24     : OPERATING SYSTEM: UNIX
25     : SOFTWARE: In house
26     : CURRENT APPLICATION DATA:
27     : APPLICATION NUMBER: US/07/942,245
28     : FILING DATE: 03-SEP-1992
29     : CLASSIFICATION: 530
30     : TELECOMMUNICATION INFORMATION:
31     : TELEPHONE: (202) 293-7060
32     : TELEFAX: (202) 293-7660
33     : TELEX: 6491103
34     : INFORMATION FOR SEQ ID NO: 312:
35     : SEQUENCE CHARACTERISTICS:
36     : LENGTH: 26 amino acids
37     : TYPE: amino acid
38     : TOPOLOGY: linear
39     : MOLECULE TYPE: peptide
40     : US-07-942-245-312

```

Query Match	32.9%	Score 46;	DB 1;	Length 26;
Best Local Similarity	45.0%	Pred. No. 0.4;		
Matches	9;	Conservative	2;	Mismatches 9;
				Indels

QY	3	EAKHRENVPGHERMGRGRTS	22
		:       :	
Db	5	ELKPGKGTPGHEKKGTSS	24

## RESULT 6

```

US-09-134-001C-4091
: Sequence 4091, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
:   APPLICANT: Lynn Doucelte-Stamm et al
:   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
:   TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
:   FILE REFERENCE: GTC-007
:   CURRENT APPLICATION NUMBER: US/09/134,001C
:   CURRENT FILING DATE: 1998-08-13
:   PRIOR APPLICATION NUMBER: US 60/064,964
:   PRIOR FILING DATE: 1997-11-08
:   PRIOR APPLICATION NUMBER: US 60/055,779
:   PRIOR FILING DATE: 1997-08-14
:   NUMBER OF SEQ ID NOS: 5674
:   SEQ ID NO 4091
:   LENGTH: 537
:   TYPE: PRNT
:   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4091

```

Query Match	32.5%	Score 45.5;	DB 4;	Length 537;
Best Local Similarity	55.0%;	Pred. No. 21;		
Matches 11; Conservative	3;	Mismatches 5;	Indels 1;	Gaps 1;

QY 9 NVPGH-ERMGRGRTSSKELA 27  
|:| | | :| | |  
Db 186 NLPGHVEAMGVSQTSTLLTA 205

```

1      RESULT 7
2      US-07-942-245-314
3      ; Sequence 314, Application US/07942245
4      ; Patent No. 5639641
5      ; General INFORMATION:
6      ; APPLICANT: PEDERSEN, Jan T.
7      ; APPLICANT: SEARLE, Stephen M.J.
8      ; APPLICANT: REES, Anthony R.
9      ; APPLICANT: ROGUSKA, Michael A.
10     ; APPLICANT: GUILD, Braydon C.
11     ; TITLE OF INVENTION: SURFACE RESIDUE VENERERING OF RODENT
12     ; TITLE OF INVENTION: ANTIBODIES
13     ; NUMBER OF SEQUENCES: 522
14     ; CORRESPONDENCE ADDRESS:
15     ; ADDRESSEE: Suthner, Mion, Zinn, Macpeak & Seas
16     ; STREET: 2100 Pennsylvania Avenue, N.W.
17     ; CITY: Washington
18     ; STATE: D.C.
19     ; COUNTRY: United States
20     ; ZIP: 20037-3202
21     ; COMPUTER READABLE FORM:
22     ; MEDIUM TYPE: Floppy disk
23     ; COMPUTER: HP 9000/700 Workstation
24     ; OPERATING SYSTEM: UNIX
25     ; SOFTWARE: In house
26     ; CURRENT APPLICATION DATA:
27     ; APPLICATION NUMBER: US/07/942,245
28     ; FILING DATE: 09-SEP-1992
29     ; CLASSIFICATION: 530
30     ; TELECOMMUNICATION INFORMATION:
31     ; TELEPHONE: (202) 293-7060
32     ; TELEFAX: (202) 293-7860
33     ; TELEX: 6491103
34     ; INFORMATION FOR SEQ ID NO: 314:
35     ; SEQUENCE CHARACTERISTICS:
36     ; LENGTH: 26 amino acids
37     ; TYPE: amino acid
38     ; TOPOLOGY: linear
39     ; MOLECULE TYPE: peptide
40     ;
41     ; US-07-942-245-314

```

Query Match	31.48;	Score 44;	DB 1;	Length 26;
Best Local Similarity	40.08;	Pred. No. 0.86;		

Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 3 EAKHRENVGHERMGRGRTS 22  
| : : |||| : ||  
Db 5 EVRPGKGTGPHKKGTSSTS 24

RESULT 8  
US-07-942-245-315  
; Sequence 315, Application US/07942245  
; Patent No. 5639641  
; GENERAL INFORMATION:  
; APPLICANT: PEDERSEN, Jan T.  
; APPLICANT: SEARLE, Stephen M.J.  
; APPLICANT: REES, Anthony R.  
; APPLICANT: ROGUSKA, Michael A.  
; APPLICANT: GUILD, Braydon C.  
; TITLE OF INVENTION: SURFACE RESIDUE VENERING OF RODENT  
; NUMBER OF SEQUENCES: 522  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: HP 9000/700 Workstation  
; OPERATING SYSTEM: UNIX  
; SOFTWARE: in house  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/942,245  
; FILING DATE: 09-SEP-1992  
; CLASSIFICATION: 530  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 293-7060  
; TELEFAX: (202) 293-7860  
; TELEX: 6491103  
; INFORMATION FOR SEQ ID NO: 315:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-942-245-315

Query Match 31.4%; Score 44; DB 1; Length 26;  
Best Local Similarity 45.0%; Pred. No. 0.86;  
Matches 9; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 3 EAKHRENVGHERMGRGRTS 22  
| : : |||| : ||  
Db 5 ELKPGKGTGPHKKGTSSTS 24

RESULT 9  
US-07-942-245-349  
; Sequence 349, Application US/07942245  
; Patent No. 5639641  
; GENERAL INFORMATION:  
; APPLICANT: PEDERSEN, Jan T.  
; APPLICANT: SEARLE, Stephen M.J.  
; APPLICANT: REES, Anthony R.  
; APPLICANT: ROGUSKA, Michael A.  
; APPLICANT: GUILD, Braydon C.  
; TITLE OF INVENTION: SURFACE RESIDUE VENERING OF RODENT  
; NUMBER OF SEQUENCES: 522  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue, N.W.

City: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: HP 9000/700 Workstation  
OPERATING SYSTEM: UNIX  
SOFTWARE: in house  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/942,245  
FILING DATE: 09-SEP-1992  
CLASSIFICATION: 530  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 349:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-942-245-349

Query Match 31.4%; Score 44; DB 1; Length 26;  
Best Local Similarity 45.0%; Pred. No. 0.86;  
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 3 EAKHRENVGHERMGRGRTS 22  
| : : |||| : ||  
Db 5 ELKPGKGTGPHKKGTSSTS 24

RESULT 10  
US-09-026-482B-2  
; Sequence 2, Application US/09026482B  
; Patent No. 6143538  
; GENERAL INFORMATION:  
; APPLICANT: REISER, STEVEN E.  
; APPLICANT: SOMMERVILLE, CHRIS  
; TITLE OF INVENTION: ACYL-CoA REDUCTASE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL A. GOTTLIEB, AGCTT  
; ADDRESSEE: DEPARTMENT OF ENERGY  
; ADDRESS: GC-62 (FORSTL) MS-6F-067  
; STREET: 1000 INDEPENDENCE AVE. S.W.  
; CITY: WASHINGTON, D.C.  
; ZIP: 20585  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORDPERFECT 6.22  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/026,482B  
; FILING DATE: 02/19/98  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ALMAN, JOY  
; REGISTRATION NUMBER: 40486  
; REFERENCE/DOCKET NUMBER: S-87814  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 630-252-2179  
; TELEFAX: 630-252-2779  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-026-482B-2

Query Match 31.4%; Score 44; DB 4; Length 295;  
Best Local Similarity 39.4%; Pred. No. 18;  
Matches 13; Conservative 4; Mismatches 10; Indels 6; Gaps 1;  
Qy 1 RLEAKHRENVGHERM-----GGRTSSKELA 27  
Db 4 KLEALFRENWKGVALITGASSGIGLTIARRIA 36

RESULT 11  
US-09-134-001C-4012  
Sequence 4012, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4012  
LENGTH: 617  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4012

Query Match 31.4%; Score 44; DB 4; Length 617;  
Best Local Similarity 50.0%; Pred. No. 44;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
Qy 3 EAKHRENVGHERM 16  
Db 92 EDKHEKDVPGHKL 105

RESULT 12  
US-07-942-245-311  
Sequence 311, Application US/07942245  
Patent No. 5639641  
GENERAL INFORMATION:  
APPLICANT: PEDERSEN, Jan T.  
APPLICANT: SEARLE, Stephen M.J.  
APPLICANT: REES, Anthony R.  
APPLICANT: ROGUSKA, Michael A.  
APPLICANT: GUILD, Braydon C.  
TITLE OF INVENTION: SURFACE RESIDUE VENERING OF RODENT  
NUMBER OF SEQUENCES: 522  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: HP 9000/700 workstation  
OPERATING SYSTEM: UNIX  
SOFTWARE: in house  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/942,245  
FILING DATE: 09-SEP-1992  
CLASSIFICATION: 530  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860

TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 311:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-942-245-311

Query Match 30.7%; Score 43; DB 1; Length 26;  
Best Local Similarity 40.0%; Pred. No. 13;  
Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
Qy 3 EAKHRENVGHEMGRTS 22  
Db 5 ELRPGKTPGHEKKGTS 24

RESULT 13  
US-08-581-528A-7  
Sequence 7, Application US/08581528A  
Patent No. 5986058  
GENERAL INFORMATION:  
APPLICANT: Lee, Se-Jin  
APPLICANT: Huynh, Thanh  
TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-7  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/581,528A  
FILING DATE: 03-Sept-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/089,670  
FILING DATE: 09-JUL-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lisa A. Halle, Ph.D.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/081001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IMMEDIATE SOURCE:  
CLONE: GDP-7  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..119  
US-08-581-528A-7

Query Match 30.7%; Score 43; DB 2; Length 119;  
Best Local Similarity 60.0%; Pred. No. 83;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 12 GHERMGRGRTSSKEL 26

Db 9 GHGRRGRSRSKSL 23

## RESULT 14

PCT-US94-07799-7

Sequence 7 Application PC/TUS9407799  
APPLICANT: THE JOHN HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-7  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Jubas & Lubitz  
STREET: 1880 Century Park East, Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07799  
FILING DATE: 08-JUL-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: TUMARKIN, LISA A., PH.D.  
REGISTRATION NUMBER: P-38,347  
REFERENCE/DOCKET NUMBER: PD-2348  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IMMEDIATE SOURCE:  
CLONE: GDF-7  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..119  
PCT-US94-07799-7  
Query Match 30.7%; Score 43; DB 5; Length 119;  
Best Local Similarity 60.0%; Pred. No. 8.3;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 12 GHERMGRTSSKSL 26  
Db 9 GHGRRGRSRSKSL 23

RESULT 15  
US-08-581-528A-6  
Sequence 6, Application US/08581528A  
Patent No. 5986058  
GENERAL INFORMATION:  
APPLICANT: Lee, Se-Jin  
APPLICANT: Huynh, Thanh  
TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-7  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/581,528A  
FILING DATE: 03-Sept-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/089,670  
FILING DATE: 09-JUL-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lisa A. Haile, Ph.D.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/081001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5099  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-581-528A-6

Query Match 30.7%; Score 43; DB 2; Length 161;  
Best Local Similarity 60.0%; Pred. No. 12;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 12 GHERMGRTSSKSL 26  
Db 51 GHGRRGRSRSKSL 65

Search completed: October 29, 2002, 14:06:26  
Job time : 27 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 13:55:48 ; Search time 31 Seconds  
(without alignments)  
116.057 Million cell updates/sec

Title: US-09-674-913a-2  
Perfect score: 140  
Sequence: 1 RLEAKHRENVDPGHERMGRGRTSKELA 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
tal number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_101002:\*

- 1: /SID2/gcgdata/geneseq/genesep-emb1/AA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/genesep-emb1/AA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/genesep-emb1/AA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/genesep-emb1/AA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/genesep-emb1/AA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/genesep-emb1/AA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/genesep-emb1/AA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/genesep-emb1/AA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/genesep-emb1/AA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/genesep-emb1/AA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/genesep-emb1/AA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/genesep-emb1/AA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/genesep-emb1/AA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/genesep-emb1/AA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/genesep-emb1/AA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/genesep-emb1/AA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/genesep-emb1/AA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/genesep-emb1/AA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/genesep-emb1/AA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/genesep-emb1/AA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/genesep-emb1/AA2001.DAT:\*
- 23: /SID2/gcgdata/geneseq/genesep-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	140	100.0	27	21	AAV56392
2	140	100.0	28	19	AAV21521
3	140	100.0	48	22	AAV99217
4	109	77.9	25	19	AAV20219
5	109	77.9	300	18	AAV16643
6	99	70.7	19	21	AAV56391
7	97	69.3	28	19	AAV21522
8	93	66.4	17	21	AAV56393
9	82.5	58.9	19	22	AAV99216
10	63	45.0	14	22	AAV99215

11	58	41.4	12	21	AAV56394	Mutant beta-APP pr
12	51	36.4	57	22	AAU46738	Proionibacterium
13	51	36.4	57	22	AAU66292	Proionibacterium
14	50.5	36.1	491	22	ABR59130	Drosophila melanog
15	50	35.7	26	15	AAK52350	Mouse heavy chain
16	50	35.7	72	23	ABR53619	Lactococcus lactis
17	50	35.7	180	20	AAU45481	Proionibacterium
18	49	35.0	140	20	AAV59810	Human normal ovar
19	49	35.0	193	22	ABR68806	Drosophila melanog
20	48	34.3	31	22	ABR38968	Peptide #6674 enco
21	48	34.3	31	22	ABR23939	Protein #5938 enco
22	48	34.3	31	22	AAV59619	Human brain expres
23	48	34.3	31	22	AAV72192	Human bone marrow
24	48	34.3	31	22	AAV19485	Peptide #5919 enco
25	48	34.3	31	22	AAV32452	Peptide #6489 enco
26	48	34.3	31	23	ABG42006	Human peptide enco
27	48	34.3	912	23	ABP61501	Human NF-KB activa
28	47.5	33.9	1006	23	ABP61502	Human NF-KB activa
29	47.5	33.9	226	22	ABG08110	Novel human diagno
30	47.5	33.9	332	9	AAV80534	Deacetoxycephalosp
31	47.5	33.9	332	22	AAV69572	Cephalosporium acr
32	47	33.6	10	18	AAV18633	Peptide resulting
33	47	33.6	10	19	AAV20152	Human beta-APP Imm
34	47	33.6	10	22	AAV99214	AMY1 peptide. Uni
35	47	33.6	46	21	AAV57802	Arabidopsis thalia
36	47	33.6	117	21	AAV12574	Zea mays protein f
37	46.5	33.2	56	22	AAV56187	Proionibacterium
38	46.5	33.2	281	21	AAV29741	Arabidopsis thalia
39	46.5	33.2	288	22	AAV79320	Corynebacterium g1
40	46.5	33.2	296	22	AAV69927	Corynebacterium prote
41	46.5	33.2	296	22	AAV79319	Corynebacterium g1
42	46	32.9	26	15	AAV52341	Mouse heavy chain
43	46	32.9	138	22	AAV42121	Proionibacterium
44	46	32.9	329	21	AAV51794	Gene 21 human sect
45	46	32.9	349	22	AAV56594	Novel protein kina

#### ALIGNMENTS

RESULT 1	AAV56392	AAV56392 standard; peptide: 27 AA.
ID	AAV56392	
XX	XX	
AC	AAV56392;	
XX	XX	
DT	15-FEB-2000 (first entry)	
XX	XX	
DE	Mutant beta-APP protein peptide sequence SEQ ID NO:2.	
XX	XX	
KM	Mutant, beta-amyloid precursor protein; beta-APP; ubiquitin-B; vaccine;	
KM	Alzheimer's disease; Down syndrome; T cellular immunity; neuroprotective;	
XX	gene therapy.	
OS	Synthetic.	
OS	Homo sapiens.	
XX	XX	
PN	WO9958564-A1.	
XX	XX	
PD	18-NOV-1999.	
XX	XX	
PF	30-APR-1999; 99WO-NO00141.	
XX	XX	
PR	08-MAY-1998; 98NO-0002098.	
XX	XX	
PA	(NHVD ) NORSK HYDRO AS.	
XX	XX	
PI	Gaudernack G, Eriksen JA, Moller M;	
XX	XX	
DR	WPI: 2000-039070/03.	
XX	XX	
PT	Frameshift mutant beta-amyloid precursor and ubiquitin-B peptides,	
	useful for treating Alzheimer's disease and Down syndrome	

XX PS- Claim 10; Page 27; 33pp; English.  
XX CC The present invention describes frameshift mutant beta-amyloid precursor  
CC peptides (beta-APP) (AAV56391 to AAV56394) and mutant ubiquitin-B  
CC (Ubi-B) (AAV56395 to AAV56400) associated with Alzheimer's disease and  
CC Down syndrome eliciting T cellular immunity. The peptides may be used as  
CC a vaccine for Alzheimer's disease and Down syndrome. The vaccination is  
CC sufficient to induce specific T-cell immunity to mutant beta-APP and/or  
CC Ubi-B peptides that are associated with Alzheimer's disease or Down  
CC syndrome. The patient may be stimulated in vivo or ex vivo with the  
CC peptides. The peptides and DNA encoding the peptides can also be used for  
CC the treatment or prophylaxis of Alzheimer's disease or the treatment of  
CC Down syndrome.  
XX SO Sequence 27 AA:  
Query Match 100.0%; Score 140; DB 21; Length 27;  
Best Local Similarity 100.0%; Pred. No. 6.6e-16;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 RLEAKHRENVGHERMGRGRTSKELA 27  
1 RLEAKHRENVGHERMGRGRTSKELA 27  
Db  
RESULT 2  
AAV21521  
ID AAV21521 standard; Protein: 28 AA.  
XX AC AAV21521;  
XX DT 22-JUL-1999 (first entry)  
XX DE Human beta-APP exon 9 protein fragment.  
XX KW Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;  
KW frameshift mutation; age-related disease; neurodegenerative disorder;  
KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;  
KW Huntington's disease; multiple sclerosis; alcoholic liver disease;  
KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;  
KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;  
KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;  
KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;  
KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;  
KW high mobility group protein-C; neuroendocrine specific protein A.  
XX Homo sapiens.  
XX WO9845322-A2.  
XX PD 15-OCT-1998.  
XX PF 02-APR-1998; 98WO-IB00705.  
XX PR 10-APR-1997; 97US-0043163.  
XX PA (UYUT-) RIJKSUNIV UTRECHT.  
XX PA (ROYA-) ROYAL NETHERLANDS ACADEMIES OF SCI.  
XX PA (UTRO-) UNIV ROTTERDAM ERASMUS.  
XX PI Burbach JPH, Grosveld FG, Van Leeuwen FW;  
XX WPI: 1998-609901/51.  
XX DR N-PSDB; AAX75771.  
XX PT Diagnosing disease by detecting frameshift mutations in RNA or  
XX PT corresponding protein mutations - used to diagnose cancer and  
XX PT neurological diseases, particularly Alzheimer's disease, and also  
XX PT for treatment and prevention with specific ribozymes or wild-type  
XX RNA  
XX PS Disclosure; Figure 20; 258pp; English.

XX CC This invention describes a novel method for the diagnosis of a disease  
CC caused by, or associated with, an RNA molecule that has a frameshift  
CC mutation. The method is used to diagnose age-related diseases, especially  
CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's  
CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,  
CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II  
CC and many others listed) or susceptibility to these disorders. The method  
CC allows a definitive diagnosis of Alzheimer's disease in living patients,  
CC at an early stage. It is based on the observation that disease may be  
CC caused by mutations in RNA rather than DNA. The invention describes the  
CC use of neuronal system RNA molecules, specifically proteins including  
CC beta-amyloid precursor protein (beta-APP), the microtubule associated  
CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule  
CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,  
CC neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic  
CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma  
CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group  
CC protein-C (HMGP-C) and neuroendocrine specific protein A.  
XX SO Sequence 28 AA:  
Query Match 100.0%; Score 140; DB 19; Length 28;  
Best Local Similarity 100.0%; Pred. No. 6.9e-16;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 RLEAKHRENVGHERMGRGRTSKELA 27  
2 RLEAKHRENVGHERMGRGRTSKELA 28  
Db  
RESULT 3  
AAB99217  
ID AAB99217 standard; Protein: 48 AA.  
XX AC AAB99217;  
XX DT 06-SEP-2001 (first entry)  
XX DE Beta-amyloid precursor protein, APP, protein fragment.  
XX KW Disease detection; age-related disease; Alzheimer's disease;  
KW Down's syndrome; cancer; neurodegenerative disease; Parkinson's disease;  
KW amyotrophic lateral sclerosis; Huntington's disease;  
KW spinocerebellar ataxia-3; multiple sclerosis; diabetes mellitus type II;  
KW degenerative disease; cardiovascular disease; rheumatoid arthritis;  
KW beta-amyloid precursor protein; APP.  
XX Unidentified.  
XX OS  
XX PN WO200140804-A2.  
XX PD 07-JUN-2001.  
XX PF 04-DEC-2000; 2000WO-NI00887.  
XX PR 03-DEC-1999; 99EP-0204140.  
XX PA (NEW-) KONINK NEDERLANDSE AKADEMIJE VAN WETENSCHAPPE.  
XX PA Hol EM, Van Leeuwen FW;  
XX WPI: 2001-397965/42.  
XX DR  
XX PT Detecting a disease or pathological condition, associated with  
XX PT secretion of aberrant protein e.g. age-related diseases, by determining  
XX PT level of aberrant protein and/or its detectable part in body fluid  
XX PT and/or tissue  
XX PS Examples; Fig 2; 36pp; English.  
XX CC The present invention relates to a method for detecting a disease or  
XX CC pathological condition associated with molecular misreading of coding



CC sequences in the genome and/or associated with the clearance of aberrant  
 CC protein. The method comprises providing a sample of a body fluid and/or  
 CC tissue of a patient and determining a level of the aberrant protein  
 CC and/or its detectable part or the ratio between the two in the body fluid  
 CC and/or tissue. The method enables diagnosing the disease before the  
 CC patient exhibits clinical symptoms. The method is useful for diagnosing  
 CC age-related diseases, preferably Alzheimer's disease, Down's syndrome or  
 CC other age related diseases e.g. cancer; neurodegenerative diseases, such  
 CC as frontal lobe dementia, progressive supranuclear palsy; and other  
 CC diseases with abundant tau-positive filamentous lesions, Parkinson's  
 CC disease, amyotrophic lateral sclerosis, Huntington's disease,  
 CC spinocerebellar ataxia-3, multiple sclerosis; other inclusion body  
 CC diseases associated with ubiquitin, diabetes mellitus type II and other  
 CC degenerative diseases such as cardiovascular diseases and rheumatoid  
 CC arthritis. A number of other diseases which can be diagnosed are given in  
 CC the specification. The present sequence is a fragment of the beta-amyloid  
 CC precursor protein (APP), which was used in the examples from the present  
 CC invention.

XX Sequence 48 AA;  
 QY Query Match 100.0%; Score 140; DB 22; Length 48;  
 D5 Best Local Similarity 100.0%; Pred. No. 1.3e-15;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4  
 ID AAY20219 standard; Protein: 25 AA.  
 XX AAY20219;

22-JUL-1999 (first entry)

DE Human beta-amyloid precursor protein mutant fragment 44.

XX Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;  
 KW frameshift mutation; age-related disease; neurodegenerative disorder;  
 KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;  
 KW Huntington's disease; multiple sclerosis; alcoholic liver disease;  
 KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;  
 KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;  
 KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;  
 KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;  
 KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;  
 KW high mobility group protein-C; neuroendocrine specific protein A.

XX Synthetic.

OS Homo sapiens.

XX MO9845322-A2.

XX 15-OCT-1998.

XX 02-APR-1998; 98WO-IB00705.

XX 10-APR-1997; 97US-0043163.

XX (UYUT-) RITKSUNIV UMRECHT.

XX (ROYA-) ROYAL NETHERLANDS ACADEMIES & SCI.

XX (UYKO-) UNIV ROTTERDAM ERASMUS.

XX Burdach JPH, Grosveld FG, Van Leeuwen FW;

XX WPI: 1998-609901/51.

XX N-PSDB: AAX75753.

XX Diagnosing disease by detecting frameshift mutations in RNA or  
 PT corresponding protein mutations - used to diagnose cancer and

PT neurological diseases, particularly Alzheimer's disease, and also  
 PT for treatment and prevention with specific ribozymes or wild-type  
 PT RNA

PS Disclosure; Figure 2; 258bp; English.

XX This invention describes a novel method for the diagnosis of a disease  
 CC caused by, or associated with, an RNA molecule that has a frameshift  
 CC mutation. The method is used to diagnose age-related diseases, especially  
 CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's  
 CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,  
 CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II  
 CC and many others listed) or susceptibility to these disorders. The method  
 CC allows a definitive diagnosis of Alzheimer's disease in living patients,  
 CC at an early stage. It is based on the observation that disease may be  
 CC caused by mutations in RNA rather than DNA. The invention describes the  
 CC use of neuronal system RNA molecules, specifically proteins including  
 CC beta-amyloid precursor protein (beta-APP), the microtubule associated  
 CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule  
 CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,  
 CC neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic  
 CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma  
 CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group  
 CC protein-C (HMGP-C) and neuroendocrine specific protein A.

SQ Sequence 25 AA;

QY Query Match 77.9%; Score 109; DB 19; Length 25;  
 D5 Best Local Similarity 100.0%; Pred. No. 7.8e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 7 RENVPGHERMGRTSSKELA 27  
 5 RENVPGHERMGRTSSKELA 25

# \*RESULT 5

ID AAM18643 standard; Protein: 300 AA.

XX AAM18643;

22-JUL-1997 (first entry)

DE Fragmented human amyloid A4 gene +1 frameshift mutant product.

XX Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;  
 KW detection; antibody; probe; cancer; neoplasia; neurodegenerative;  
 KW Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;  
 KW Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;  
 KW PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;  
 KW cardiovascular; rheumatoid arthritis.

XX Homo sapiens.

XX Key Location/Qualifiers

XX MISC-difference 1..300 /note- "X corresponds to a stop codon in the

XX Peptide 179..188 accompanying DNA file, AAT69789"

XX /note- "antigenic peptide used for antibody

XX MO9712992-A2.

XX 10-APR-1997.

XX 02-OCT-1996; 96WO-IB01106.

XX 11-JAN-1996; 96US-0009832.

XX 02-OCT-1995; 95GB-0020080.

XX (ROYA-) ROYAL NETHERLANDS ACADEMIES & SCI.

PA (UYRO-) UNIV ROTTERDAM ERASMUS.  
 PA (UYUT-) UNIV STATE UTRECHT.  
 PI Burbach JPH, Grosveld FG, Van Leeuwen FW.  
 XX WPI: 1997-226235/20.  
 DR N-PSDB: AAY69789.  
 PT Use of mutant genes having frame:shift mutation(s) - for developing  
 PT prods. for the diagnosis, prevention and treatment of associated  
 PT diseases, e.g. cancer or neuro:degenerative disease  
 XX  
 PS Claim 22: Fig 2, 123pp: English.  
 XX  
 CC AAY18642 and AAY18643 are +2 and +1 frameshift mutations, respectively,  
 CC of a sequence comprising fragments of the coding sequence of the human  
 CC amyloid A4 gene corresponding to nucleotides (nt) 147-566 followed by nt  
 CC 1047-1526 of the wild-type amyloid A4 gene. These regions of the gene  
 CC contain GAGAG motifs. Frameshift mutants of the tau, ubiquitin,  
 CC apolipoprotein E, microtubule-associated protein 2 (MAP-2),  
 CC neurofilament subunit L, M and H and amyloid A4 genes are claimed. All  
 CC these genes share a common GAGAN motif (N- A, G, C or T), which is the  
 CC site of common GA dinucleotide deletion(s) that cause neurodegenerative  
 CC disorders. Antigenic peptides used for the production of antibodies, and  
 CC small nucleic acid sequences derived from frameshift mutants are used in  
 CC the diagnosis, prevention and treatment of cancer and neurodegenerative  
 CC diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's  
 CC syndrome, frontal lobe dementia (Pick's disease), progressive  
 CC supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's  
 CC disease, multiple sclerosis, and other degenerative diseases such as  
 CC cardiovascular disease and rheumatoid arthritis.  
 CC  
 SQ Sequence 300 AA:  
 QY  
 DE 7 RENVGHERMGRGRTSSKELA 27  
 DB 168 RENVGHERMGRGRTSSKELA 188  
 ID  
 RESULT 6  
 ID AAY56391 standard; peptide; 19 AA.  
 XX  
 XX AAY56391;  
 XX  
 XX 15-FEB-2000 (first entry)  
 XX  
 DE Mutant beta-APP protein peptide sequence SEQ ID NO:1.  
 XX  
 XX Mutant; beta-amyloid precursor protein; beta-APP; ubiquitin-B; vaccine;  
 KM Alzheimer's disease; Down syndrome; T cellular immunity; neuroprotective;  
 KM gene therapy.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO958564-A1.  
 XX  
 PD 18-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-MO00141.  
 XX  
 PR 08-MAY-1998; 98NO-0002098.  
 XX  
 PA (NHYD ) NORSK HYDRO AS.  
 XX  
 PI Gaudernack G, Eriksen JA, Moller M;  
 XX  
 DR WPI: 2000-039070/03.

XX  
 PT Frameshift mutant beta-amyloid precursor and ubiquitin-B peptides,  
 PT useful for treating Alzheimer's disease and Down syndrome -  
 XX  
 XX Claim 10; Page 27; 33pp: English.  
 PS  
 CC The present invention describes frameshift mutant beta-amyloid precursor  
 CC peptides (beta-APP) (AAY56391 to AAY56394) and mutant ubiquitin-B  
 CC (Ubl-B) (AAY56395 to AAY56400) associated with Alzheimer's disease and  
 CC Down syndrome eliciting T cellular immunity. The peptides may be used as  
 CC a vaccine for Alzheimer's disease and Down syndrome. The vaccination is  
 CC sufficient to induce specific T-cell immunity to mutant beta-APP and/or  
 CC Ubl-B peptides that are associated with Alzheimer's disease or Down  
 CC syndrome. The patient may be stimulated in vivo or ex vivo with the  
 CC peptides. The peptides and DNA encoding the peptides can also be used for  
 CC the treatment or prophylaxis of Alzheimer's disease or the treatment of  
 CC Down syndrome.  
 CC  
 SQ Sequence 19 AA:  
 QY  
 DE 9 NVPGERMGRGRTSSKELA 27  
 DB 1 NVPGERMGRGRTSSKELA 19  
 ID  
 RESULT 7  
 ID AAY21522  
 XX  
 XX AAY21522 standard; Protein; 28 AA.  
 AC  
 XX  
 DT 22-JUL-1999 (first entry)  
 XX  
 DE Human beta-APP exon 10 protein fragment.  
 XX  
 KM Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;  
 KM frameshift mutation; age-related disease; neurodegenerative disorder;  
 KM Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;  
 KM Huntington's disease; multiple sclerosis; alcoholic liver disease;  
 KM diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;  
 KM ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;  
 KM neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;  
 KM glial fibrillary acidic protein; GFAP; p53; semaphorin IIT; HSPF-1;  
 KM bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; MSP-A;  
 KM high mobility group protein-C; neuroendocrine specific protein A.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO9845322-A2.  
 XX  
 PD 15-OCT-1998.  
 XX  
 PF 02-APR-1998; 98WO-IB00705.  
 XX  
 PR 10-APR-1997; 97US-0043163.  
 XX  
 PA (UYUT-) RIJKSUNIV UTRECHT.  
 PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.  
 PA (UYRO-) UNIV ROTTERDAM ERASMUS.  
 XX  
 PI Burbach JPH, Grosveld FG, Van Leeuwen FW.  
 XX  
 DR WPI: 1998-609901/51.  
 DR N-PSDB: AAY75772.  
 XX  
 PT Diagnosing disease by detecting frameshift mutations in RNA or  
 PT corresponding protein mutations - used to diagnose cancer and  
 PT neurological diseases, particularly Alzheimer's disease, and also

PT		for treatment and prevention with specific ribozymes or wild-type
PI	RNA	
PS		Disclosure; Figure 20; 258bp; English.
XX		
CC		This invention describes a novel method for the diagnosis of a disease
CC		caused by, or associated with, an RNA molecule that has a frameshift
CC		mutation. The method is used to diagnose age-related diseases, especially
CC		cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
CC		disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
CC		multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
CC		and many others listed) or susceptibility to these disorders. The method
CC		allows a definitive diagnosis of Alzheimer's disease in living patients,
CC		at an early stage. It is based on the observation that disease may be
CC		caused by mutations in RNA rather than DNA. The invention describes the
CC		use of neuronal system RNA molecules, specifically proteins including
CC		beta-amyloid precursor protein (beta-ApP), the microtubule associated
CC		proteins Tau and Btg Tau, ubiquitin B, apolipoprotein E, microtubule
CC		associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
CC		neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic
CC		protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
CC		2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
CC		protein-C (HMGP-C) and neuroendocrine specific protein A.
CC		
SO	Sequence	28 AA:
-	Query Match	69.3%; Score 97; DB 19; Length 28;
-	Best Local Similarity	77.8%; Pred. No. 8.4e-09;
Matches	21; Conservative	1; Mismatches 5; Indels 0; Gaps 0
OY	1 RLEAKHRENYPGHERMGRTSSKELA 27	
	:	
Db	2 RLEAKHREMSOYWIRMGRTSSKELA 28	
RESULT 8		
AAY56393		
ID	AAY56393 standard; peptide; 17 AA.	
XX		
AC	AAY56393;	
XX		
DE	15-FEB-2000 (first entry)	
XX		
DN		
XX	Mutant beta-APP protein peptide sequence SEQ ID NO:3.	
XX		
KM	Mutant; beta-amyloid precursor protein; beta-APP; ubiquitin-B; vaccine;	
KW	Alzheimer's disease; Down syndrome; T cellular immunity; neuroprotective;	
KX	gene therapy.	
OS	Synthetic.	
OS	Homo sapiens.	
XX		
PN	WO950564-A1.	
PD	18-NOV-1999.	
XX		
PF	30-APR-1999; 99MO-N000141.	
XX		
PR	08-MAY-1998; 98NO-0002098.	
XX		
PA	(NH2D ) NORSK HYDRO AS.	
PI	Gaudernack G, Eriksen JA, Moeller M;	
DR	WPI: 2000-039070/03.	
XX		
PT	Frameshift mutant beta-amyloid precursor and ubiquitin-B peptides,	
XX	useful for treating Alzheimer's disease and Down syndrome -	
PS	Claim 10; Page 27; 33pp; English.	
CC	The present invention describes frameshift mutant beta-amyloid precursor	
CC	peptides (beta-APP) (AAY56391 to AAY56394) and mutant ubiquitin-B	

Query Match	Best Local Similarity	Score 93:	DB 21:	Length 17:
Matches 17: Conservative	100.0%:	Pred. No. 2.2e-08:	Mismatches 0:	Indels 0: Gaps 0:
1 RLEAKHRENVPGHERMG 17				
1 RLEAKHRENVPGHERMG 17				
RESULT 9				
AAB99216				
ID AAB99216 standard: peptide: 19 AA.				
XX AAB99216:				
XX 06-SEP-2001 (first entry)				
XX AMY6 peptide.				
XX AMY6, disease detection; age-related disease; Alzheimer's disease;				
XX Down's syndrome; cancer; neurodegenerative disease; Parkinson's disease;				
XX amyotrophic lateral sclerosis; Huntington's disease;				
XX spinocerebellar ataxia-7; multiple sclerosis; diabetes mellitus type II;				
XX degenerative disease; cardiovascular disease; rheumatoid arthritis;				
KW beta-amyloid precursor protein; APP.				
XX Unidentified.				
XX WO200140804-A2.				
XX 07-JUN-2001.				
XX 04-DEC-2000; 2000WO-NL00087.				
XX 03-DEC-1999; 99EP-0204140.				
XX (NEWA-) KONINK NEDERLANDSE AKAD VAN WETENSCHAPPE.				
XX Hol EM, Van Leeuwen FW;				
XX WPI: 2001-397965/42.				
XX Detecting a disease or pathological condition, associated with				
XX secretion of aberrant protein e.g. age-related diseases, by determining				
XX level of aberrant protein and/or its detectable part in body fluid				
XX and/or tissue				
XX Disclosure: Page 24; 36pp; English.				
XX The present invention relates to a method for detecting a disease or				
XX pathological condition associated with molecular misreading of coding				
XX sequences in the genome and/or associated with the clearance of aberrant				
XX protein. The method comprises providing a sample of a body fluid and/or				
XX tissue of a patient and determining a level of the aberrant protein				
XX and/or its detectable part or the ratio between the two in the body fluid				
XX and/or tissue. The method enables diagnosing the disease before the				
XX patient exhibits clinical symptoms. The method is useful for diagnosing				
XX age-related diseases, preferably Alzheimer's disease, Down's syndrome or				
XX other age related diseases e.g. cancer; neurodegenerative diseases, such				
XX as frontal lobe dementia, progressive supranuclear palsy, and other				
XX diseases with abundant tau-positive filamentous lesions, Parkinson's				

CC disease, amyotrophic lateral sclerosis, Huntington's disease,  
CC spinocerebellar ataxia-3, multiple sclerosis; other inclusion body  
CC diseases associated with ubiquitin, diabetes mellitus type II and other  
CC degenerative diseases such as cardiovascular diseases and rheumatoid  
CC arthritis. A number of other diseases which can be diagnosed are given in  
CC the specification. The present sequence is AMY 6 peptide, which was  
CC coupled to thyroglobulin by glutaraldehyde, and used to immunise  
CC rabbits for antibody production in the present invention. The AMY 6  
CC peptide is a C-terminal fragment of the full-length beta-amyloid  
CC precursor protein (APP+1; mature APP and its signal peptide) protein.  
XX  
SQ Sequence 19 AA:  
XX  
Query Match 58.9%; Score 82.5; DB 22; Length 19;  
Best Local Similarity 94.7%; Pred. No. 1.3e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 9 NYPGHRMGRGRTSSKELA 27  
2 NVP-HERMGRGRTSSKELA 19  
IIIIIIIIIIIIIIIIIIII  
RESULT 10  
AAB99215  
ID AAB99215 standard; peptide; 14 AA.  
XX  
AC AAB99215;  
XX  
DT 06-SEP-2001 (first entry)  
XX  
DE AMY5 peptide.  
XX  
AMY5; disease detection; age-related disease; Alzheimer's disease;  
KM Down's syndrome; cancer; neurodegenerative disease; Parkinson's disease;  
KM amyotrophic lateral sclerosis; Huntington's disease;  
KM spinocerebellar ataxia-3; multiple sclerosis; diabetes mellitus type II;  
KM degenerative disease; cardiovascular disease; rheumatoid arthritis;  
KM beta-amyloid precursor protein; APP.  
XX  
OS Unidentified.  
XX  
PN WO200140804-A2.  
XX  
PD 07-JUN-2001.  
XX  
PF 04-DEC-2000; 2000WO-NL00887.  
XX  
PR 03-DEC-1999; 99EP-0204140.  
XX  
PA (NEWM-) KONINK NEDERLANDSE AKAD VAN WETENSCHAPPE.  
XX  
PI HOL EM, Van Leeuwen FW;  
XX  
DR WPI; 2001-397965/42.  
XX  
PT Detecting a disease or pathological condition, associated with  
PT secretion of aberrant protein e.g. age-related diseases, by determining  
PT level of aberrant protein and/or its detectable part in body fluid  
PT and/or tissue  
XX  
PS Disclosure: Page 24; 36pp; English.  
XX  
The present invention relates to a method for detecting a disease or  
CC pathological condition associated with molecular misreading of coding  
CC sequences in the genome and/or associated with the clearance of aberrant  
CC protein. The method comprises providing a sample of a body fluid and/or  
CC tissue of a patient and determining a level of the aberrant protein  
CC and/or its detectable part or the ratio between the two in the body fluid  
CC and/or tissue. The method enables diagnosing the disease before the  
CC patient exhibits clinical symptoms. The method is useful for diagnosing  
CC age-related diseases, preferably Alzheimer's disease, Down's syndrome or  
CC other age related diseases e.g. cancer; neurodegenerative diseases, such  
CC as frontal lobe dementia, progressive supranuclear palsy; and other

CC diseases with abundant tau-positive filamentous lesions, Parkinson's  
CC disease, amyotrophic lateral sclerosis, Huntington's disease,  
CC spinocerebellar ataxia-3, multiple sclerosis; other inclusion body  
CC diseases associated with ubiquitin, diabetes mellitus type II and other  
CC degenerative diseases such as cardiovascular diseases and rheumatoid  
CC arthritis. A number of other diseases which can be diagnosed are given in  
CC the specification. The present sequence is AMY 5 peptide, which was  
CC coupled to thyroglobulin by glutaraldehyde, and used to immunise  
CC rabbits for antibody production in the present invention. The AMY 5  
CC peptide is a C-terminal fragment of the full-length beta-amyloid  
CC precursor protein (APP+1; mature APP and its signal peptide) protein.  
XX  
SQ Sequence 14 AA:  
XX  
Query Match 45.0%; Score 63; DB 22; Length 14;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 RMGRGRTSSKELA 27  
2 RMGRGRTSSKELA 14  
IIIIIIIIIIIIIIIIIIII  
RESULT 11  
AAV56394  
ID AAV56394 standard; peptide; 12 AA.  
XX  
AC AAV56394;  
XX  
DT 15-FEB-2000 (first entry)  
XX  
DE Mutant beta-APP protein peptide sequence SEQ ID NO:4.  
XX  
KM Mutant; beta-amyloid precursor protein; beta-APP; ubiquitin-B; vaccine;  
KM Alzheimer's disease; Down syndrome; T cellular immunity; neuroprotective;  
KM gene therapy.  
XX  
OS Synthetic.  
XX  
OS Homo sapiens.  
XX  
PN WO958564-A1.  
XX  
PD 18-NOV-1999.  
XX  
PF 30-APR-1999; 99WO-NO00141.  
XX  
PR 08-MAY-1998; 98NO-0002098.  
XX  
PA (NHYD ) NORSK HYDRO AS.  
XX  
PI Gadernack G, Eriksen JA, Moller M;  
XX  
DR WPI; 2000-039070/03.  
XX  
PT Frameshift mutant beta-amyloid precursor and ubiquitin-B peptides,  
PT useful for treating Alzheimer's disease and Down syndrome  
XX  
PS Claim 10; Page 27; 33pp; English.  
XX  
The present invention describes frameshift mutant beta-amyloid precursor  
CC peptides (beta-APP) (AAV56391 to AAV56394) and mutant ubiquitin-B  
CC (Ubi-B) (AAV56395 to AAV56400) associated with Alzheimer's disease and  
CC Down syndrome eliciting T cellular immunity. The peptides may be used as  
CC a vaccine for Alzheimer's disease and Down syndrome. The vaccination is  
CC sufficient to induce specific T-cell immunity to mutant beta-APP and/or  
CC Ubi-B peptides that are associated with Alzheimer's disease or Down  
CC syndrome. The patient may be stimulated in vivo or ex vivo with the  
CC peptides. The peptides and DNA encoding the peptides can also be used for  
CC the treatment or prophylaxis of Alzheimer's disease or the treatment of  
CC Down syndrome.  
XX  
SQ Sequence 12 AA;

Query Match 41.4%; Score 58; DB 21; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.0085;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 MGRGRTSSKELA 27  
| | | | | | | | | |  
DB 1 MGRGRTSSKELA 12

## RESULT 12

AAU46738  
ID AAU46738 standard; Protein; 57 AA.

AC AAU46738;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #7634.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71.

DR N-PSDB; AAS59535.

PT Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

XX Example 1; SEQ ID No 7933; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
XX polypeptides. The proteins and their associated DNA sequences are used in  
XX the treatment, prevention and diagnosis of medical conditions caused by  
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
XX P. acnes is also involved in infections of bone, joints and the central  
XX nervous system, however it is particularly involved in the inflammatory  
XX lesions associated with acne vulgaris. A method for detecting the  
XX presence or absence of P. acnes in a patient comprises contacting a  
XX sample with a binding agent that binds to the proteins of the invention  
XX and determining the amount of bound protein in the sample. The  
XX polypeptides may be used as antigens in the production of antibodies  
XX specific for P. acnes proteins. These antibodies can be used to  
XX downregulate expression and activity of P. acnes polypeptides and  
XX therefore treat P. acnes infections. The antibodies may also be used as  
XX diagnostic agents for determining P. acnes presence, for example, by  
XX enzyme linked immunosorbent assay (ELISA).  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 57 AA;

Query Match 36.4%; Score 51; DB 22; Length 57;  
Best Local Similarity 81.8%; Pred. No. 0.73;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 VPGHRMGGR 20  
| | | | | | | | | |  
DB 47 VPGHRMGGR 57

## RESULT 13

AAU66292  
ID AAU66292 standard; Protein; 57 AA.

AC AAU66292;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #27188.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71.

DR N-PSDB; AAS59726.

PT Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

XX Example 1; SEQ ID No 27487; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
XX polypeptides. The proteins and their associated DNA sequences are used in  
XX the treatment, prevention and diagnosis of medical conditions caused by  
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
XX P. acnes is also involved in infections of bone, joints and the central  
XX nervous system, however it is particularly involved in the inflammatory  
XX lesions associated with acne vulgaris. A method for detecting the  
XX presence or absence of P. acnes in a patient comprises contacting a  
XX sample with a binding agent that binds to the proteins of the invention  
XX and determining the amount of bound protein in the sample. The  
XX polypeptides may be used as antigens in the production of antibodies  
XX specific for P. acnes proteins. These antibodies can be used to  
XX downregulate expression and activity of P. acnes polypeptides and  
XX therefore treat P. acnes infections. The antibodies may also be used as  
XX diagnostic agents for determining P. acnes presence, for example, by  
XX enzyme linked immunosorbent assay (ELISA).  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 57 AA;

Query Match 36.4%; Score 51; DB 22; Length 57;  
 Best Local Similarity 81.8%; Pred. No. 0.73;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 VPGHERMGRGR 20  
 |||||  
 DB 47 VPGHRRTRGRGR 57

# RESULT 14

ABB59130 standard; Protein; 491 AA.

AC ABB59130;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 4182.

XX Drosophila: developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR N-PSDB; ABL03233.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

PS Disclosure; SEQ ID NO 4182; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB1616176-AB130511), expressed DNA sequences (AB101840-AB1616175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.

CC Sequence 491 AA:

Query Match 36.1%; Score 50.5; DB 22; Length 491;

Best Local Similarity 48.0%; Pred. No. 11;

Matches 12; Conservative 3; Mismatches 3; Indels 7; Gaps 1;

QY 10 VPGHERMGRGR-----KELA 27  
 : ||||| : |||||

DB 171 LPSHERLIGRSLSQPTIELTKELA 195

# RESULT 15

AAR52350

ID AAR52350 standard; Peptide; 26 AA.

AC AAR52350;  
 XX 30-SEP-1996 (first entry)

DE Mouse heavy chain surface patch MUSTIGHADN.

XX antibody; humanised; murine; human; heavy chain; light; variable; framework region; complementarity determining region; reshaping; modelling; surface residue; modify.

OS Mus sp.

PN EP592106-A1.

PD 13-APR-1994.

PF 07-SEP-1993; 93EP-0307051.

PR 09-SEP-1992; 92US-0942245.

PA (PEDE/) PEDERSEN J T.

PA (IMMU-) IMMUNOGEN INC.

PI Guild BC, Pedersen JT, Rees AR, Roguska MA, Searle SMJ;

DR WPI: 1994-120230/15.

PT Method of resurfacing of rodent antibodies to produce humanised antibody forms - for producing non-human antibodies with improved therapeutic efficiency by presenting human surface on V-region

PS Example 1; Page 19; 230pp; English.

CC Modification of a rodent antibody or fragment by resurfacing in order to produce a humanised rodent antibody can be determined by calculating homology between murine and human antibody surfaces. In order to test the resurfacing approach of the invention, three humanisation experiments were set up. (1) traditional loop grafting; (2) resurfacing approach using most similar chain; and (3) resurfacing approach using human sequences with most similar surface residues. AAR52243-447 are the surface residue patterns in mouse heavy chain antibody variable regions. These "patches" were used in the third method, where rodent light and heavy chains were matched and the most similar human sequence found independently only over the surface residues indicated in AAR52030-67.

CC Sequence 26 AA:

Query Match 35.7%; Score 50; DB 15; Length 26;

Best Local Similarity 45.0%; Pred. No. 0.43;

Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 EAKHRENVPGHERMGRGRS 22  
 | : : ||||| : ||

DB 5 EVRPGKGTPEHEKGRSSTS 24

Search completed: October 29, 2002, 13:59:23  
 Job time : 32 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 13:55:48 : Search time 43 Seconds  
(without alignments)  
60.363 Million cell updates/sec

Title: US-09-674-913a-2

Perfect score: 140

Sequence: 1 RUEAKHRENVGHERMGRGRTSKELA 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Optimal number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54.5	38.9	157	2	H82989 conserved hypothet
2	50	35.7	72	2	H86664 hypothetical prote
3	50	35.7	229	2	H82043 guanylate kinase V
4	49.5	35.4	424	2	T39524 hypothetical prote
5	49.5	35.4	2140	2	T18543 probable cell-adhe
6	49	35.0	829	2	AH2282 hypothetical prote
7	48	34.3	244	2	UC7801 denitrific cell-ass
8	48	34.3	962	2	AB1817 translation initia
9	48	34.3	962	2	C81060 translation initia
10	47.5	33.9	332	2	A29711 deacetoxycephalosp
11	46.5	33.2	382	2	T35706 probable aspartate
12	46.5	33.2	769	2	A35848 competence regulat
13	46.5	33.2	769	2	B69604 two-component sens
14	46	32.9	302	2	AB3573 oligopeptide trans
15	46	32.9	342	2	A97325 homolog of cell di
16	46	32.9	485	2	T35663 probable transmem
17	46	32.9	614	2	F96791 hypothetical prote
18	46	32.9	650	2	T32897 hypothetical prote
19	46	32.9	898	2	H87481 ribonuclease, Rne/
20	46	32.9	1064	2	S52687 serine/threonine-s
21	46	32.9	1420	1	A44361 amiloride-sensitiv
22	45	32.1	155	1	IGB02 SRY protein homolo
23	45	32.1	179	1	IGB02 insulin-like growt
24	45	32.1	179	2	S04858 insulin-like growt
25	45	32.1	240	2	H90778 DNA-binding protei
26	45	32.1	242	2	C85640 hypothetical prote
27	45	32.1	403	2	H87569 hypothetical prote
28	45	32.1	1245	2	S51255 probable membrane
29	45	32.1	1575	2	T18545 lysobactin synthet

30	44.5	31.8	414	2	AG1997 ribulose-bisphosph
31	44.5	31.8	609	1	KSASL1 lactase (EC 1.10.3
32	44	31.4	144	2	T49706 hypothetical prote
33	44	31.4	215	2	T49743 probable rrm-type
34	44	31.4	306	2	F84405 ferric enterobacti
35	44	31.4	312	2	H89884 hypothetical prote
36	44	31.4	326	2	T40353 probable ribosomal
37	44	31.4	407	2	G90453 electron transfer
38	44	31.4	532	2	T49467 related to CopI-in
39	44	31.4	586	1	E69314 replication licens
40	44	31.4	852	2	A34373 histidine-rich cal
41	44	31.4	858	2	T09344 probable phosphol
42	44	31.4	946	2	A96748 hypothetical prote
43	44	31.4	1490	2	S72351 nonstructural poly
44	43.5	31.1	206	2	E83113 30S ribosomal prot
45	43.5	31.1	498	2	T34592 DEAD-box RNA helic

## ALIGNMENTS

## RESULT 1

H82989

conserved hypothetical protein PA5246 [imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

H82989

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

Adman, S.; Yuan, Y.; Brody, L.D.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

..; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: AB2950; MUID:20437337; PMID:10584043

A:Accession: H82989

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-157 <STO>

A:Cross-references: GB:AE004937; GB:AE004091; NID:9951553; PIDN:AG08631.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA5246

Query Match 38.9%; Score 54.5; DB 2; Length 157;

Best Local Similarity 40.0%; Pred. No. 1.1;

Matches 14; Conservative 2; Mismatches 0; Indels 19; Gaps 2;

QY 5 KHEENVGHERM-----GGRGT 21

DB 85 RHRE-LPGHERMARLSKLTIDLRVDYLRPGRGT 118

## RESULT 2

AB6664

hypothetical protein ydbc [imported] - Lactococcus lactis subsp. lactis (strain IL140

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C:Accession: AB6664

R:Boilotin, A.;

Wincker, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Eh

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis

A:Reference number: AB6625; MUID:21235186; PMID:11337471

A:Accession: AB6664

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-72 <STO>

A:Cross-references: GB:AE005176; PID:912723178; PIDN:AAK04411.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: ydbc

Query Match 35.7%; Score 50; DB 2; Length 72;

Best Local Similarity 60.0%; Pred. No. 2.2;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;





Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 8 ENVPGHERMGRG 19  
| | | | | : | : |  
Db 18 ETVPGHQRLERG 29

## RESULT 8

translation initiation factor IF-2 NMA1897 [similarity] - Neisseria meningitidis (strain C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: A81817  
R:Perkhill, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, J.; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, A.; et al. 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: A81817  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-962 <PAR>  
A:Cross-references: GB:AL162757; GB:AL157959; NID:97380371; PIDN:CAB85118.1; PID:9738053  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: infB; NMA1897  
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homolog  
F:465-574/Domain: translation elongation factor Tu homology <ETU>

Query Match 34.3%; Score 48; DB 2; Length 962;  
Best Local Similarity 40.9%; Pred. No. 59;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 3 EAKHRENVGHERMGRGRTSSK 24  
| | | | | : | : |  
Db 195 EPREKAKPKHERNGKGDAAK 216

## RESULT 9

translation initiation factor IF-2 NMB1643 [imported] - Neisseria meningitidis (strain C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: C81060  
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Hart, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; et al. 2000  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; et al. 2000  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: C81060  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-962 <TET>  
A:Cross-references: GB:AE002514; GB:AE002098; NID:97226886; PIDN:AAFA1992.1; PID:9722689  
C:Genetics:  
A:Experimental source: serogroup B, strain MC58  
A:Gene: NMB1643  
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homolog  
F:465-574/Domain: translation elongation factor Tu homology <ETU>

Query Match 34.3%; Score 48; DB 2; Length 962;  
Best Local Similarity 40.9%; Pred. No. 59;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 3 EAKHRENVGHERMGRGRTSSK 24  
| | | | | : | : |  
Db 195 EPREKAKPKHERNGKGDAAK 216

RESULT 10  
A29711

deacetoxycephalosporin C synthetase - fungus (Acremonium sp.)  
N:Alternate names: expandase; hydroxylase  
C:Species: Acremonium sp.  
C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 19-May-2000  
C:Accession: A29711; A41864  
R:Samson, S.M.; Dotzler, J.E.; Slisz, M.L.; Becker, G.W.; Van Frank, R.M.; Veal, L.E.  
Bio/Technology 5, 1207-1214, 1987  
A:Title: Cloning and expression of the fungal expandase/hydroxylase gene involved in  
A:Reference number: A29711  
A:Accession: A29711  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-332 <SAM>

A:Note: the source is designated as Cephalosporium acremonium  
R:Butler, S.; Velasco, J.; Fernandez, F.J.; Martin, J.F.  
J. Bacteriol. 174, 3056-3064, 1992  
A:Title: The cefG gene of Cephalosporium acremonium is linked to the cefEF gene and e  
A:Reference number: A41864; MUID:92234966; PMID:1569032  
A:Accession: A41864  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-76, 'L' <GUT>  
A:Experimental source: strain C10  
A:Note: sequence extracted from NCBI backbone (NCBI:104773, NCBI:97574); this ORF 1  
A:Note: the source is designated as Cephalosporium acremonium  
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 33.9%; Score 47.5; DB 2; Length 332;  
Best Local Similarity 55.0%; Pred. No. 24;  
Matches 11; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY 5 KHRENVGHERMGRGRTSS 23  
| | | | | : | : |  
Db 243 KHRENVGHERMGRGRTSS 262

## RESULT 11

probable aspartate transaminase (EC 2.6.1.1) SC7H1.11 [similarity] - Streptomyces coe  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 20-Jun-2000  
C:Accession: T35706  
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, January 1998  
A:Reference number: Z21548  
A:Accession: T35706  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-382 <MUR>  
A:Cross-references: EMBL:AL021411; PIDN:CA16198.1; GSPDB:GNO0070; SCORDB:SC7H1.11  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCORDB:SC7H1.11  
C:Superfamily: aspartate transaminase  
C:Keywords: aminotransferase

Query Match 33.2%; Score 46.5; DB 2; Length 382;  
Best Local Similarity 46.2%; Pred. No. 39;  
Matches 12; Conservative 1; Mismatches 6; Indels 7; Gaps 1;

QY 3 EAKHRENVGHERMGRG-----RGRT 21  
| | | | | : | : |  
Db 197 EAYHRLAYPGHEPVSALESRLRGRT 222

## RESULT 12

A35848  
competence regulatory protein Comp - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999  
C:Accession: A35848; D38111  
R:Weinrauch, Y.; Penchev, R.; Dubnau, E.; Smith, I.; Dubnau, D.  
Genes Dev. 4, 860-872, 1990

A>Title: A *Bacillus subtilis* regulatory gene product for genetic competence and sporulation  
A/Reference number: A35848; MUID:90337321; PMID:2116363  
A/Accession: A35848  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-769 <ME1>  
A/Cross-references: GB:M22856; NID:q142697; PIDN:AAA2319.1; PID:q142700  
R:Welntrauch, Y.; Masdek, T.; Kunst, F.; Dubnau, D.  
J. Bacteriol. 173, 5685-5693, 1991  
A>Title: Sequence and properties of *comO*, a new competence regulatory gene of *Bacillus subtilis*  
A/Reference number: A38111; MUID:91358358; PMID:1715859  
A/Accession: D38111  
A:Molecule type: DNA  
A:Residues: 1-16 <ME2>  
A/Cross-references: GB:M71283; NID:q142712; PIDN:AAA2324.1; PID:q142715  
C/Genetics:  
A:Gene: *comP*  
A:Start codon: ATG  
C/Keywords: autophosphorylation; phosphoprotein; transmembrane protein

Query Match 33.2%; Score 46.5; DB 2; Length 769;  
Best Local Similarity 44.4%; Pred. No. 79;  
Matches 12; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

QY 1 RLEAKHRENVGHERMGR-TSSKEL 26  
: | : | | | : | | : | : | : |  
DB 649 KLVAFEDQERVPFHRLNTGRTASLDL 675

RESULT 13  
B69604  
Two-component sensor histidine kinase involved in early competent comp - *Bacillus subtilis*  
C/Species: *Bacillus subtilis*  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C/Accession: B69604  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte, C.; Bron, S.; Bouilliet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc A.; Ehlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetille, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A/Reference number: A69580; MUID:98044033; PMID:9384377  
A/Accession: B69604  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-769 <KUN>  
A/Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15157.1; PID:e1184246; A:Experimental source: strain 168  
C/Genetics:  
A:Gene: *comp*

Query Match 33.2%; Score 46.5; DB 2; Length 769;  
Best Local Similarity 44.4%; Pred. No. 79;  
Matches 12; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

QY 1 RLEAKHRENVGHERMGR-TSSKEL 26  
: | : | | | : | | : | : | : |  
DB 649 KLVAFEDQERVPFHRLNTGRTASLDL 675

RESULT 14  
AB3573  
Oligopeptide transport ATP-binding protein oppD BMEI10507 [imported] - *Brucella melitensis*  
C/Species: *Brucella melitensis*

C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
C/Accession: AB3573  
R:DeVechio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Muje, C.; Los, T.; Ivanov, J.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A/Reference number: AD3552; PMID:11756688  
A/Accession: AB3573  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-302 <KUR>  
A/Cross-references: GB:AE008918; PIDN:AAL53749.1; PID:q17984675; GSPDB:GN00191  
A:Experimental source: strain 16M  
C/Genetics:  
A:Gene: BMEI10507  
A:Map position: II  
C/Superfamily: inner membrane protein malK; ATP-binding cassette homology

Query Match 32.9%; Score 46; DB 2; Length 302;  
Best Local Similarity 40.0%; Pred. No. 37;  
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 8 ENVPHERMGRRTSKELA 27  
: | : | | | : | : | : | : |  
DB 151 ETRAHENSGRGARAKAIA 170

RESULT 15  
A97325  
Homolog of cell division GTPase FtsZ, diverged [imported] - *Clostridium acetobutylicum*  
C/Species: *Clostridium acetobutylicum*  
C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C/Accession: A97325  
R:Nozling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R. J. Bacteriol. 183, 4823-4838, 2001  
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A/Reference number: A96900; MUID:21359325; PMID:21359325  
A/Accession: A97325  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-342 <KUR>  
A/Cross-references: GB:AE001437; PIDN:AAK81388.1; PID:g15026550; GSPDB:GN00168  
A:Experimental source: *Clostridium acetobutylicum* ATCC824  
C/Genetics:  
A:Gene: CAC3459

Query Match 32.9%; Score 46; DB 2; Length 342;  
Best Local Similarity 56.2%; Pred. No. 42;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 11 PGHERMGRRTSKEL 26  
: | : | | | : | : | : | : |  
DB 57 PGTDSGRDRKSKEM 72

Search completed: October 29, 2002, 14:00:27  
Job time : 44 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 13:55:48 ; Search time 34 Seconds

(without alignments)  
32.937 Million cell updates/sec

Title: US-09-674-913A-2

Perfect score: 140  
Sequence: 1 RLEAKHRENVGHEMRGRTSKELA 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

tal number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	39.3	576	1	LE12_RALSO
2	47.5	33.9	332	1	EXPA_CBPAC
3	46.5	33.2	769	1	COMP_BACSU
4	46	32.9	1064	1	KINI_YEAST
5	46	32.9	1420	1	APX_XENIA
6	45	32.1	155	1	IGF2_BOVIN
7	45	32.1	179	1	IGF2_SHEEP
8	44.5	31.8	414	1	RCA_ANASP
9	44.5	31.8	609	1	LACT_EMENT
10	44	31.4	220	1	SRT_CANFA
11	44	31.4	539	1	DOP2_DROME
12	44	31.4	852	1	SRCH_RABIT
13	44	31.4	858	1	PDGI_ARATH
14	44	31.4	1093	1	RNT1_NEUCR
15	43.5	31.1	206	1	RS4_PSEAF
16	43	30.7	151	1	GDF7_MOUSE
17	43	30.7	181	1	IGF2_PIG
18	43	30.7	201	1	YH23_SCHPO
19	43	30.7	206	1	CASI_PIG
20	43	30.7	208	1	SRY_PIG
21	43	30.7	285	1	CASB_TRTIV
22	43	30.7	855	1	CLOC_MOUSE
23	43	30.7	1944	1	CHD3_HUMAN
24	42.5	30.4	415	1	RCA_ANASC
25	42.5	30.4	758	1	VKGC_BOVIN
26	42	30.0	95	1	VE5B_BPV4
27	42	30.0	355	1	RECA_RHOCA
28	42	30.0	356	1	GLN2_SOYBN
29	42	30.0	356	1	GLN2_SOYBN
30	42	30.0	406	1	RECA_PARDE
31	42	30.0	687	1	YB86_AQUAE
32	42	30.0	751	1	TRA_HAEIN
33	42	30.0	770	1	A4_SATSC
					A4_HUMAN
					P03067 homo sapien

34	42	30.0	770	1	A4_MOUSE	P12023 mus musculus
35	42	30.0	770	1	A4_RAT	P08592 ratu mus norv
36	42	30.0	846	1	CLOC_HUMAN	O15156 homo sapien
37	42	30.0	915	1	TBP1_NEIGO	O01996 neisseria g
38	42	30.0	1150	1	C9EA_BACTA	O92n19 bacillus th
39	41	29.3	123	1	RS13_MYCBO	P45813 mycobacteri
40	41	29.3	200	1	YNU2_SHITL	P29771 shigella fl
41	41	29.3	217	1	RR3_PINTH	P41635 pinus thunb
42	41	29.3	218	1	RR3_PICAB	O62951 picea abies
43	41	29.3	255	1	LIV3_SALTY	P30293 salmonella
44	41	29.3	315	1	GLN3_LUPAN	P14636 lupinus ang
45	41	29.3	331	1	RECA_ACHLA	P29225 acholeplasm

## ALIGNMENTS

RESULT 1  
LE12\_RALSO STANDARD: PRT: 576 AA.

AC 08XS25:  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE 2-isopropylmalate synthase 2 (EC 4.1.3.12) (Alpha-isopropylmalate synthase 2) (Alpha-IPM synthetase 2).  
GN LEUA2 OR RSP0322 OR RS05445.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OX Ralstonia.  
RN NCBI\_TaxID=305;  
RP STRAIN=GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Arlat M., Billault A., Brotlier P., Camus J.C., Cattolico L., Chandel M., Choise N., Claudel-Renard C., Cunne S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T., Sigler P., Tneault P., Whalen M., Winkler P., Levy M., Weissenbach J., Boucher C.A.;  
RL "Genome sequence of the plant pathogen Ralstonia solanacearum"; Nature 415:497-502(2002).

-1- FUNCTION: Catalyzes the condensation of the acetyl group of acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).  
-1- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA -> acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.  
-1- PATHWAY: Leucine biosynthesis; first step.  
-1- SUBUNIT: Homotetramer (BY similarity).  
-1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE SYNTHASE FAMILY. LEUA 2 SUBFAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL: AL646078; CAD17473.1; -  
InterPro: IPR002034; AIPM/HcIt-synth.  
InterPro: IPR000891; HMG-L-like.

DR Pfam: PF00682; HMG-L-like; 1.  
DR TIGRfam: TIGR00970; leua-yeast; 1.  
DR PROSITE: PS00815; AIPM\_HOMOCITR\_SYNTH\_1; 1.  
DR PROSITE: PS00816; AIPM\_HOMOCITR\_SYNTH\_2; 1.  
KW Leucine biosynthesis; Lyase; Plasmid; Complete proteome.  
SEQUENCE 576 AA: 63149 MW; BCB0A9A66BA332B CRC64;

Query Match 39.3%; Score 55; DB 1; Length 576;



```

CC -----
DR EMBL: X54010; CAA37957.1; -
DR EMBL: Z93932; CAB07903.1; -
DR EMBL: Z99120; CAB15157.1; -
DR EMBL: M22856; AAA22319.1; -
DR EMBL: M71283; AAA22324.1; -
DR PIR: A35848; A35848.
DR Subtilist; BG10380; comp.
DR InterPro: IPR003594; AtPp1ind ATPase.
DR InterPro: IPR004359; His_Kin_slg.
DR Pfam: PF02518; HATPase_c; 1.
DR SMART: SM00387; HATPase_c; 1.
DR PROSITE: PS0109; His_Kin_1.
KW Sensority transduction; Phosphorylation; Transferase; Kinase;
KW Transmembrane; Complete proteome.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 1 33 POTENTIAL.
FT DOMAIN 34 113 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 114 134 POTENTIAL.
FT DOMAIN 135 144 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 145 167 POTENTIAL.
FT DOMAIN 168 235 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 236 257 POTENTIAL.
FT DOMAIN 258 272 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 273 295 POTENTIAL.
FT DOMAIN 296 299 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 300 323 POTENTIAL.
FT DOMAIN 324 337 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 338 357 POTENTIAL.
FT DOMAIN 358 361 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 362 383 POTENTIAL.
FT DOMAIN 384 769 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 571 769 HISTIDINE KINASE.
FT MOD_RES 456 456 S -> C (IN REF. 1).
FT CONFLICT 604 604 S -> Y (IN REF. 1).
FT CONFLICT 610 610 E -> G (IN REF. 1).
FT CONFLICT 628 628 D -> PV (IN REF. 1).
FT CONFLICT 636 637 OL -> PV (IN REF. 1).
SQ SEQUENCE 769 AA; 89316 MW; E5C8ACE278DE5B4 CRC64;

Query Match 33.2%; Score 46.5; DB 1; Length 769;
Best Local Similarity 44.4%; Pred. No. 32;
Matches 12; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

OY 1 RLEAKHRENVGHERMGRGR-TSSKEL 26
DB 649 KLVAEQERVPHIRLMTGRTASLDL 675

RESULT 4
KIN1_YEAST STANDARD; PRT; 1064 AA.
ID KIN1_YEAST
AC P13185; 004606;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Protein kinase KIN1 (EC 2.7.1.1).
GN KIN1 OR YDR122W OR YD9727.17.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87317589; PubMed=2957690;
RA Levin D.E., Hammond C.I., Ralston R.O., Bishop J.M.;
RA "Two yeast genes that encode unusual protein kinases.";
RA Proc. Natl. Acad. Sci. U.S.A. 84:6035-6039(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=S288C / AB972;
RA Murphy L., Shore L., Harris D., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;

```

```

RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS PROBABLY A SERINE/THREONINE PROTEIN
CC KINASE.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC NIMI SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M69017; AAA34722.1; -
DR EMBL: Z48758; CAA88675.1; -
DR PIR: S42438; S42438.
DR HSSP: Q63450; 1A06.
DR SGD: S0002529; KIN1.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001772; Kinase_Cterm.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF02149; KAI_1.
DR PRODOM: PD000001; Euk_pkinase; 1.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_SF; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 120 398 PROTEIN KINASE.
FT NP_BIND 126 134 ATP (BY SIMILARITY).
FT BINDING 149 149 ATP (BY SIMILARITY).
FT ACT_SITE 269 269 S -> R (IN REF. 1).
FT CONFLICT 25 25 S -> R (IN REF. 1).
FT CONFLICT 453 453 T -> H (IN REF. 1).
FT CONFLICT 455 455 V -> G (IN REF. 1).
FT CONFLICT 718 718 A -> R (IN REF. 1).
FT CONFLICT 718 718 NI -> IN (IN REF. 1).
FT CONFLICT 920 921 NI -> A (IN REF. 1).
FT CONFLICT 976 976 SI -> T (IN REF. 1).
FT CONFLICT 979 980 SI -> NS (IN REF. 1).
FT CONFLICT 984 985 KT -> NS (IN REF. 1).
SQ SEQUENCE 1064 AA; 120070 MW; E4104B84A8E45C36 CRC64;

Query Match 32.9%; Score 46; DB 1; Length 1064;
Best Local Similarity 41.7%; Pred. No. 53;
Matches 10; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

OY 2 LEAKHRENVGHERMGRGRTSSKE 25
DB 83 LEGKRENAPEKNTTSOSRVSSO 106

RESULT 5
APX_XENLA STANDARD; PRT; 1420 AA.
ID APX_XENLA
AC Q01613;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Apical protein (ApX).
GN APX.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Ovary;
RA MEDLINE=93107151; PubMed=1334959;
RA Staub O., Verrey F., Kleyman T.R., Benos D.J., Rossier B.C.,
RA Kraehenbuhl J.-P.;

```



DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-OCT-1989 (Rel. 12, last sequence update)  
 DT 01-OCT-1996 (Rel. 34, last annotation update)  
 DE Insulin-like growth factor II precursor (IGF-II).  
 GN IGF2.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OC NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA MEDLINE=89345107; PubMed=2762134;  
 RT O'Mahoney J.V., Adams T.E.;  
 RL "Nucleotide sequence of an ovine insulin-like growth factor-II cDNA."; Nucleic Acids Res. 17:5392-5392(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA MEDLINE=90356421; PubMed=2388846;  
 RT Brown W.M., Dziegielewska K.M., Foreman R.C., Saunders N.R.;  
 RL "The nucleotide and deduced amino acid sequences of insulin-like growth factor II cDNAs from adult bovine and fetal sheep liver."; Nucleic Acids Res. 18:4614-4614(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COPworth; TISSUE=Liver;  
 RA MEDLINE=93250051; PubMed=8485157;  
 RT Demmer J., Hill D.F., Petersen G.B.;  
 RL "Characterization of two sheep insulin-like growth factor II cDNAs with different 5'-untranslated regions."; Biochim. Biophys. Acta 1173:79-80(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Ohlsen S.M., Wong E.A.;  
 RL Submitted (SEP-1990) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 25-91.  
 RA MEDLINE=89136887; PubMed=2537174;  
 RT Francis G.L., McNeill K.A., Wallace J.C., Ballard F.J., Owens P.C.;  
 RL "Sheep insulin-like growth factors I and II: sequences, activities and assays."; Endocrinology 124:1173-1183(1989).  
 RN [6]  
 RP SEQUENCE OF 25-58.  
 RA MEDLINE=89323215; PubMed=2752053;  
 RT Hey A.W., Browne C.A., Simpson R.J., Thorburn G.D.;  
 RL "Simultaneous isolation of insulin-like growth factors I and II from adult sheep serum."; Biochim. Biophys. Acta 997:27-35(1989).  
 CC -I- FUNCTION: THE INSULIN-LIKE GROWTH FACTORS POSSESS GROWTH-PROMOTING ACTIVITY IN VITRO. THEY ARE POTENT MITOGENS FOR CULTURED CELLS. IGF-II IS INFLUENCED BY PLACENTAL LACTOGEN AND MAY PLAY A ROLE IN FETAL DEVELOPMENT.  
 CC -I- SUBCELLULAR LOCATION: Secreted.  
 CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; 000668; AAB60626.1; -;  
 DR EMBL; 000666; AAB60626.1; JOINED.  
 DR EMBL; 000667; AAB60626.1; JOINED.  
 DR EMBL; X15248; CAA33324.1; -;  
 DR EMBL; X53554; CAA37621.1; -;  
 DR EMBL; M89788; AAA31548.1; -;

DR EMBL; M89789; AAA31549.1; -;  
 DR EMBL; X55638; CAA39163.1; -;  
 DR PIR; S04858; S04858.  
 DR PIR; S08567; S08567.  
 DR PIR; S10984; S10984.  
 DR PIR; S20731; S20731.  
 DR PIR; S32557; S32557.  
 DR HSSP; P01344; IGR2.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00276; INSULINA.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF. 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Mitogen; Growth factor; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 91 INSULIN-LIKE GROWTH FACTOR II.  
 FT DOMAIN 25 52 B.  
 FT DOMAIN 53 64 C.  
 FT DOMAIN 65 85 D.  
 FT DOMAIN 86 91 A.  
 FT PROPE 92 179 E PEPTIDE.  
 FT DISULFD 33 71 BY SIMILARITY.  
 FT DISULFD 45 84 BY SIMILARITY.  
 FT DISULFD 70 75 BY SIMILARITY.  
 FT CONFLICT 46 47 GD -> DG (IN REF. 5).  
 SQ SEQUENCE 179 AA; 19616 MW; 7B369AE57FE2E4378 CRC64;  
 Query Match 32.1%; Score 45; DB 1; Length 179;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
 Oy 7 RENVPHERMGRGRTSKEL 26  
 Db 127 RGLPAFLARRGRRLAKEL 146  
 RESULT 8  
 ID RCA\_ANASP STANDARD; PRT; 414 AA.  
 AC P58555;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, last sequence update)  
 DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Ribulose biphosphate carboxylase activase (Rubisco activase) (Ra).  
 GN RCA OR ALR1533.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 ON NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21595285; PubMed=11759840;  
 RA Koneko T., Nakamura Y., Wolk C.P.,  
 RA Matsuda A., Iritaguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Nakazaki N., Kohara M., Matsunoto M., Matsuno A., Muraki A.,  
 RA Yasuda M., Shimpo S., Sugimoto M., Takasawa M., Yamada M.,  
 RA "Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120."; DNA Res. 8:205-213(2001).  
 CC -I- FUNCTION: ACTIVATION OF RUBISCO (RUBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE; EC 4.1.1.39) INVOLVES THE ATP-DEPENDENT CARBOXYLATION OF THE EPSILON-AMINO GROUP OF LYSINE LEADING TO A CARBAMATE STRUCTURE (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO THE RUBISCO ACTIVASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003586; BAB789.1; -.
RC ATP-binding; Complete proteome.
FW NP_BIND 37 44
SQ SEQUENCE 414 AA; 46668 MW; E8359F10D8C558DA CRC64;

Query Match 31.8%; Score 44.5; DB 1; Length 414;
Best Local Similarity 39.3%; Pred. No. 33;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

OY 9 NVPG-----HERMGRGRTSSKELA 27
      |||      | | | | | |||
Db 23 NIPGVRPVLIIIGIRKRGKGFQCELA 50

RESULT 9
LACL_EMENI STANDARD; PRT; 609 AA.
PI7489:
01-AUG-1990 (Rel. 15, Created)
01-FEB-1996 (Rel. 33, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Laccase precursor (BC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
(Urishiol oxidase) (Laccase I).
YA.
Emericella nidulans (Aspergillus nidulans).
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
NCBI_TaxID=5072;
[]]
SEQUENCE FROM N.A.
STRAIN=Winter, and FGSC 4;
MEDLINE=90287738; PubMed=2192364;
Aranayo R., Timberlake W.E.;
"Sequence and molecular structure of the Aspergillus nidulans yA
(laccase I) gene.";
Nucleic Acids Res. 18:3415-3415(1990).
RL
CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE).
CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
CC H(2)O.
CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE, THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINOCLEAR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted (potential).
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on ways
to use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52552; CAA36787.1; -.
CC PTR: S10149; KSASL.
CC DR InterPro; IPR001117; Cu-oxidase.
CC DR InterPro; IPR002355; Multic_u_oxidase2.
CC DR Pfam; PF00394; Cu-oxidase; 2.
CC DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; FALSE_NEG.
CC DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
CC DR Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
CC Glycoprotein; Repeat.
CC KW SIGNAL 1 20
CC FT CHAIN 21 609 POTENTIAL.
CC FT DOMAIN 45 141 IACCASE.
CC FT DOMAIN 270 372 PLASTOCYANIN-LIKE 1.
CC FT DOMAIN 463 602 PLASTOCYANIN-LIKE 2.
CC FT DOMAIN 79 79 PLASTOCYANIN-LIKE 3.
CC FT METAL 81 81 COPPER (TYPE 2) (PROBABLE).
CC FT METAL 81 81 COPPER (TYPE 3) (PROBABLE).

```

```

FT METAL 123 123 COPPER (TYPE 3) (PROBABLE).
FT METAL 125 125 COPPER (TYPE 3) (PROBABLE).
FT METAL 508 508 COPPER (TYPE 1) (PROBABLE).
FT METAL 511 511 COPPER (TYPE 2) (PROBABLE).
FT METAL 513 513 COPPER (TYPE 3) (PROBABLE).
FT METAL 585 585 COPPER (TYPE 3) (PROBABLE).
FT METAL 586 586 COPPER (TYPE 1) (PROBABLE).
FT METAL 587 587 COPPER (TYPE 3) (PROBABLE).
FT METAL 591 591 COPPER (TYPE 1) (PROBABLE).
FT METAL 596 596 COPPER (TYPE 1) (PROBABLE).
FT CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 609 AA; E667A42EF5909512 CRC64;

Query Match 31.8%; Score 44.5; DB 1; Length 609;
Best Local Similarity 37.9%; Pred. No. 50;
Matches 11; Conservative 2; Mismatches 5; Indels 11; Gaps 1;

QY 6 HRENVP-----GHERMGREGRTSS 23
   ||||| | : | | | | :
Db 378 HRENVPKKAIGPHDTVGIMTGGGNTSA 406

RESULT 10
SRY_CANFA STANDARD; PRT; 220 AA.
ID SRY_CANFA          STANDARD; PRT; 220 AA.
AC Q9X760.
DT 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sex-determining region Y protein (Testis-determining factor).
GN SRY.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99296061; PubMed=10369387;
RX Meyers-Wallen V.N., Schlatter D., Barr I., Lovell-Badge R., Keyzner A.;
RT "Sir-negative XX sex reversal in purebred dogs.";
RL Mol. Reprod. Dev. 53:266-273(1999).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC
CC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE
CC SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL
CC INTERCALATION IN THE MINOR GROOVE.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AF107021; RAD40225.1; -.
CC DR HSSP; C05066; IHRY.
CC DR InterPro; IPRO00910; HMG_12_box.
CC DR pfam; PF00505; HMG_box; 1.
CC DR SMART; SM00398; HMG; 1.
CC DR DNA-binding; Nuclease protein; Transcription regulation; Activator;
CC KW Sexual differentiation.
FT DNA_BIND 53 121 HMG_BOX.
SQ SEQUENCE 220 AA; 25505 MW; 7AA1AA74FBAAC28E CRC64;
```





QY 2 LEAKRENVPGHMRGRGRTSSKEIA 27  
 DB 384 LSAHHNGLARHHRMGKRFSLSRKIA 409

RESULT 12  
 SRCH\_RABIT STANDARD: PRT: 852 AA.  
 ID SRCH\_RABIT STANDARD: PRT: 852 AA.  
 AC P16230;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sarcoplasmic reticulum histidine-rich calcium-binding protein precursor.  
 GN HRC OR HCP.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 NCBI\_TaxID=9986;  
 [1]  
 SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=90036884; PubMed=2808365;  
 RA Hofmann S.L., Goldstein J.L., Orth K., Moosmaw C.R., Slaughter C.A., Brown M.S.;  
 RA "Molecular cloning of a histidine-rich Ca<sup>2+</sup>-binding protein of sarcolemmal reticulum that contains highly conserved repeated elements."  
 RT J. Biol. Chem. 264:18083-18090(1989).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE REGULATION OF CALCIUM SEQUESTRATION OR RELEASE IN THE SR OF SKELETAL AND CARDIAC MUSCLE.  
 CC -1- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN.  
 CC -1- SIMILARITY: STRONG, NO HUMAN HRC.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J05080; AAA31279.1; -  
 DR PIR: A34373; A34373.  
 DR InterPro: IPR002134; HCP.  
 DR PROSITE: PS00328; HCP; 10.  
 KW Calcium-binding; Signal; Repeat.  
 KY SIGNAL  
 CHAIN 1 27 POTENTIAL.  
 FT 28 852 SARCOPLASMIC RETICULUM HISTIDINE-RICH  
 FT MOD\_RES 28 852 CALCIUM-BINDING PROTEIN.  
 FT DOMAIN 59 100 BLOCKED.  
 FT REPEAT 59 79 2 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 80 100 1-1.  
 FT DOMAIN 199 470 10 X TANDEM REPEATS, ACIDIC.  
 FT REPEAT 199 224 2-1.  
 FT REPEAT 225 253 2-2.  
 FT REPEAT 254 282 2-3.  
 FT REPEAT 283 310 2-4.  
 FT REPEAT 311 339 2-5.  
 FT REPEAT 340 367 2-6.  
 FT REPEAT 368 395 2-7.  
 FT REPEAT 423 451 2-8.  
 FT REPEAT 452 470 2-9.  
 FT REPEAT 471 585 2-10.  
 FT DOMAIN 471 721 4 X APPROXIMATE TANDEM REPEATS.  
 FT DOMAIN 721 733 POLY-GLU.  
 FT DOMAIN 780 826 METAL-BINDING (POTENTIAL).  
 SQ SEQUENCE 852 AA; 96117 MW; A43CCB3E49B930 CRC64;  
 Query Match 31.4%; Score 44; DB 1; Length 852;  
 Best Local Similarity 45.5%; Pred. No. 83;

Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;  
 QY 5 KHRNVPGH--ERMGRGRTSSK 24  
 DB 595 EHOTEVPGHHQHRMGDTDSAE 616

RESULT 13  
 PDGL\_ARATH STANDARD: PRT: 858 AA.  
 ID PDGL\_ARATH STANDARD: PRT: 858 AA.  
 AC 09T053; 048544;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase D gamma 1 (EC 3.1.4.4) (AtPLDgamma1) (PLD gamma 1)  
 DE (Choline phosphatase) (Lipophosphodiesterase II) (Lecithinase D).  
 GN PLDGMAM1 OR AT4G11850 OR T26M18.60.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 OX [1]  
 RX SEQUENCE FROM N.A.  
 RP MEDLINE=98019191; PubMed=9353280;  
 RA Qin W., Pappan K., Wang X.;  
 RA "Molecular heterogeneity of phospholipase D (PLD). Cloning of PLDgamma and regulation of plant PLDgamma, -beta, and -alpha by polyphosphoinositides and calcium."  
 RT J. Biol. Chem. 272:28267-28273(1997).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schnell C., Wandt R., Murphy G., Volckaert G., Pohl T., Duesterhoeft A., Stiekema W., Ertlan K.-D., Terryn N., Harris B., Anstorge W., Brandt P., Griwall L.A., Rieger M., Weichselgatterner M., de Simone V., Oberwalder B., Macho R., Mueller M., Kreis M., Delsen M., Puldomenech P., Watson M., Schmidheini T., Reichert B., Portetle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Hohnel J., Zimmermann W., Medler H., Ridley P., Langham S.-A., McCullagh B., Billam L., Roben J., Vandenbussche F., Van der Schueren J., Grymonprez B., Chuang I.-J., Vandenbussche F., Braeken M., Wellens J., Voet M., Bastiens I., Aert K., Defoor E., Weitzenegeger T., Bothé G., Ramsperger U., Hilbert H., Braun M., Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W., Moollman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P., Bernier S., Hempel S., Feldpausch M., Lambers J., Van den Daele H., De Keyser A., Buysbaert C., Gielen J., Villarroel R., De Clercq R., Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S., Clark L., Doggett J., Hall S., Kay W., Leonard N., McLay K., Mayes R., Pettett A., Rajandream M.-A., Lyne M., Benes V., Reichmann S., Borkova D., Bloeker H., Scharte M., Grimm G., Loehner T.-H., Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fartmann B., Grandjean K., Munier D., Herzl A., Neumann S., Argitoy F., Clabaud G., Mendenhall A., Reibelt R., Massenet O., Guigley F., Clabaud G., Mendenhall A., Reibelt R., Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S., Chetoui F., Cooke R., Berger C., Montfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Barques M., Torres A., Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse C., Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C., Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S., Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L., Sekhon M., Murray J., Shee P., Cordes M., Abur-Eldeh J., Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J., Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L., Nelson J., Speth J., Ryan E., Andrews S., Gelsel C., Layman D., Du H., Ali J., Bernhoff A., Jones K., Drone K., Cotton M., Joshi C., Antoniou B., Zidanic W., Strong C., Sun H., Lamar B., Jordan C.,

Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martienssen R., McCombie W.R.,  
 RT "Sequence and analysis of chromosome 4 of the plant *Arabidopsis*  
 RT thaliana.";  
 RL Nature 402:769-777(1999).  
 RN [3]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE:99214468; PubMed:10198096;  
 RA Fan L., Zheng S., Cui D., Wang X.,  
 RT "Subcellular distribution and tissue expression of phospholipase  
 RT Dalpha, Dbeta, and Dgamma in *Arabidopsis*.";  
 RL Plant Physiol. 119:1371-1378(1999).  
 CC -1- FUNCTION: Hydrolyzes glycerol-phospholipids at the terminal  
 CC phosphodiesteric bond. Plays an important role in various cellular  
 CC processes, including phytohormone action, vesicular trafficking,  
 CC secretion, cytoskeletal arrangement, meiosis, tumor promotion,  
 CC pathogenesis, membrane deterioration and senescence. Can use  
 CC phosphatidylserine but prefers ethanolamine-containing lipids as  
 CC substrates.  
 CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a  
 CC phosphatidate.  
 CC -1- COFACTOR: Calcium. Requires micromolar level (PIP2-dependent).  
 CC -1- ENZYME REGULATION: Inhibited by neomycin.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated. Found  
 CC mainly associated with intracellular membranes but also with  
 CC mitochondrial membranes, nuclei and clathrin-coated vesicles. Not  
 CC found in chloroplast.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in roots and flowers,  
 CC moderately in stems, leaves and seedlings and low in siliques. Not  
 CC detected in seeds.  
 CC -1- INDUCTION: Activated by wounding, heavy metal, methyl salicylate,  
 CC osmotic and salt stresses.  
 CC -1- DOMAIN: C2 domain is a calcium-binding fold, and the binding  
 CC promotes the protein association with membranes. In PLD gamma, all  
 CC the calcium-coordinating acidic amino-acids are conserved.  
 CC -1- SIMILARITY: Belongs to the phospholipase D family. C2-PLD  
 CC subfamily.  
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
 CC -1- CAUTION: Rel.1 sequence differs from that shown due to frameshifts  
 CC in position 60, 79, 260, 300, 465 and 758.  
 CC -1- CAUTION: It is uncertain whether Met-1 or Met-11 is the initiator.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF027408; AAB87672.1; ALT\_FRAME.  
 CC EMBL: AL078606; CAB44323.1; -  
 CC EMBL: AL161532; CAB78328.1; -  
 CC InterPro: IPR001736; PLD.  
 CC InterPro: IPR001008; C2.  
 CC Pfam: PF00168; C2; 1.  
 CC Pfam: PF00614; PLDC; 2.  
 CC SMART: SM00239; C2; 1.  
 CC SMART: SM00155; PLDC; 2.  
 CC PROSITE: PSS0004; C2 DOMAIN\_2; 1.  
 CC PROSITE: PSS0035; PLD; 2.  
 CC Hydrolyase; Lipid degradation; Calcium; Repeat; Multigene family.  
 KW DOMAIN 29 147  
 FT DOMAIN 364 399  
 FT DOMAIN 704 731  
 FT ACT\_SITE 369 369  
 FT ACT\_SITE 371 371  
 FT ACT\_SITE 376 376  
 FT ACT\_SITE 709 709  
 FT ACT\_SITE 711 711

FT ACT\_SITE 716 716 POTENTIAL.  
 FT CONFLICT 114 115  
 FT CONFLICT 370 370 Q -> E (IN REF. 1).  
 FT CONFLICT 377 377 A -> S (IN REF. 1).  
 FT CONFLICT 634 635 MQ -> IE (IN REF. 1).  
 SQ SEQUENCE 858 AA; 95587 MW; 3344F9B9E3A7A73 CRC64;  
 Query Match 31.4%; Score 44; DB 1; Length 858;  
 Best Local Similarity 38.2%; Pred. No. 84;  
 Matches 13; Conservative 2; Mismatches 7; Indels 12; Gaps 2;  
 QY 3 EAKHRENVPG-HERMG-----RGRTSSK 24  
 Db 50 EAKHLPMDGFHNRILGMLSGLRKKEGKSSK 83  
 ||||| : : : : :  
 RESULT 14  
 RNTL\_NEUCR  
 ID RNTL\_NEUCR STANDARD; PRT: 1093 AA.  
 AC Q9HEH1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Regulator of nonsense transcripts 1 homolog.  
 GN 264.130.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SRRJN-74-OR23-1A;  
 RA Schulte U., Aign V., Hohnsels J., Brandt P., Fartmann B., Holland R.,  
 RA Nyakatura G., Mewes H.-W., Mannhaupt G.,  
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Eliminates the production of nonsense-containing RNAs  
 CC (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).  
 CC -1- SIMILARITY: BELONGS TO THE DNM2/NM7 HELICASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AL451022; CAC18314.1; -  
 CC InterPro: IPR003593; AAA\_Atpase.  
 CC InterPro: IPR001410; DEAD.  
 CC SMART: SM00382; AAA; 1.  
 CC SMART: SM00487; DEXDC; 1.  
 CC Hypothetical protein; Hydrolyase; ATP-binding; Zinc-finger.  
 KW ZN\_FING 111 139 C2H2-type (ATYPICAL) (POTENTIAL).  
 FT ZN\_FING 163 193 CA-TYPE (POTENTIAL).  
 FT NP\_BIND 477 484 ATP (POTENTIAL).  
 FT DOMAIN 59 62 POLY-ASP.  
 FT DOMAIN 69 73 POLY-ASP.  
 SQ SEQUENCE 1093 AA; 120087 MW; 8B0E4F0407ACE142 CRC64;  
 Query Match 31.4%; Score 44; DB 1; Length 1093;  
 Best Local Similarity 42.4%; Pred. No. 1,le=02;  
 Matches 14; Conservative 5; Mismatches 6; Indels 8; Gaps 2;  
 QY 1 RLEAKHRENVPG-----HERMGRTSSKEL 26  
 Db 528 RLTAKSREDESVSFLALHEQV-RNMTNKEL 559  
 ||||| : : : : :  
 RESULT 15  
 RS4\_PSEAE  
 ID RS4\_PSEAE STANDARD; PRT: 206 AA.

```

AC 052759;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR PA4239.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Polger K.R., Kas A., Lapid K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA *Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [2]
RP SEQUENCE OF 189-206 FROM N.A.
RC STRAIN=FRD1;
RX MEDLINE=99296583; PubMed=10368148;
RA Ma J.-F., Ochser U.A., Klotz M.G., Nanayakkara V.K., Howell M.L.,
RA Johnson Z., Posey J.E., Vasil M.L., Monaco J.J., Hassett D.J.;
RT *Bacterioferritin A modulates catalase A (Kata) activity and
RT resistance to hydrogen peroxide in Pseudomonas aeruginosa.";
RL J. Bacteriol. 181:3730-3742(1999).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 S4 RNA-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE004841; AAC07627.1; -.
CC EMBL: AF047025; AAC03115.1; -.
CC HSSP: P81288; IC05.
CC InterPro: IPR001912; Ribosomal_S4.
CC InterPro: IPR002942; S4.
CC Pfam: PF00163; Ribosomal_S4; 1.
CC Pfam: PF01479; S4; 1.
CC SMART: SM00363; S4; 1.
CC DR TIGRFAMs: TIGR01017; rpsd_bact; 1.
CC DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
CC KM Ribosomal protein; rRNA-binding; Complete proteome.
CC FT CONFLICT 189 194 SDSLAD -> NSSSVP (IN REF. 2).
CC SQ SEQUENCE 206 AA; 23277 MW; DBD85D0191723186 CRC64;

Query Match 31.1%; Score 43.5; DB 1; Length 206;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 11; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 2 LEAKHR-ENVPGRHMRGRS 22
DB 28 LDSCKAENVPQHGQGRRLS 49

```

Search completed: October 29, 2002, 13:58:30  
 Job time : 35 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 13:55:48 : Search time 79 seconds  
(without alignments)  
70.421 Million cell updates/sec

Title: US-09-674-913a-2  
Perfect score: 140  
Sequence: 1 RLEAKRENVPGHERMGRGRTSKELA 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SEPREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp Unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	39.3	576	16	Q8XSZ5 Ralstonia s
2	54.5	38.9	157	16	Q8XTU8 Q8XSZ5 pseudomonas
3	50.5	36.1	251	5	Q9S5F6 Q9S5F6 dirosophila
4	50.5	36.1	491	5	Q9YI66 Q9YI66 dirosophila
5	50	35.7	72	16	Q9CIP3 Q9CIP3 lactococcus
6	50	35.7	838	10	Q9XIX9 Q9XIX9 oryza sativ
7	49.5	35.4	424	3	Q74328 Q74328 schizosach
8	49.5	35.4	2140	3	Q02316 Q02316 lentinula e
9	49	35.0	193	5	Q9V712 Q9V712 dirosophila
10	49	35.0	829	16	Q8YQX8 Q8YQX8 anabaena sp
11	48.5	34.6	483	10	Q8SBA6 Q8SBA6 oryza sativ
12	48.5	34.6	765	2	Q9KSL1 Q9KSL1 bacillus su
13	48.5	34.6	765	2	Q9KSL1 Q9KSL1 bacillus su
14	48	34.3	244	4	Q8TF63 Q8TF63 homo sapien
15	48	34.3	864	4	Q94827 Q94827 homo sapien
16	48	34.3	930	4	Q96BS1 Q96BS1 homo sapien

17	48	34.3	962	16	Q9JYD2 Q9JYD2 neisseria m
18	48	34.3	962	16	Q9JYB5 Q9JYB5 neisseria m
19	48	34.3	1349	5	Q9NE31 Q9NE31 leishmania
20	47.5	33.9	222	10	Q9M3R3 Q9M3R3 pinus sylve
21	47.5	33.9	332	3	Q9P4T5 Q9P4T5 cephalospor
22	47	33.6	113	6	Q9N1S5 Q9N1S5 capreolus c
23	47	33.6	491	5	Q76269 Q76269 leishmania
24	47	33.6	491	5	Q76343 Q76343 leishmania
25	46.5	33.2	382	16	Q54170 Q54170 streptomyce
26	46.5	33.2	476	11	Q8R1P6 Q8R1P6 mus musculu
27	46.5	33.2	770	2	Q9LC68 Q9LC68 bacillus su
28	46	32.9	87	6	P79187 P79187 macropus gi
29	46	32.9	87	6	P79187 P79187 macropus ru
30	46	32.9	188	6	P79417 P79417 petrogale b
31	46	32.9	188	6	P79418 P79418 petrogale b
32	46	32.9	188	6	P79409 P79409 petrogale x
33	46	32.9	188	6	P79414 P79414 petrogale m
34	46	32.9	188	6	P79414 P79414 petrogale c
35	46	32.9	188	6	P79989 P79989 petrogale a
36	46	32.9	188	6	P79410 P79410 petrogale x
37	46	32.9	188	6	P79413 P79413 petrogale l
38	46	32.9	188	6	P79415 P79415 petrogale p
39	46	32.9	188	6	P79411 P79411 petrogale h
40	46	32.9	188	6	P79412 P79412 petrogale l
41	46	32.9	188	6	P79408 P79408 petrogale s
42	46	32.9	188	6	P79416 P79416 petrogale g
43	46	32.9	302	16	Q8YCM1 Q8YCM1 bruceella me
44	46	32.9	313	5	Q44967 Q44967 caenorhabdi
45	46	32.9	342	16	Q97DL5 Q97DL5 clostridium

#### ALIGNMENTS

#### RESULT 1

ID	Q8XSZ5	PRELIMINARY:	PRT:	576 AA.
AC	Q8XSZ5:			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Probable 2-isopropylmalate synthase protein (DC 4.1.3.12).			
OS	LEU2A OR RSP0322 OR RS05445.			
OS	Ralstonia solanacearum (Pseudomonas solanacearum).			
OC	Plasmid megaplasmid.			
OC	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;			
OC	Ralstonia.			
OX	NCBI_TaxID=305;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RX	STRAIN-GM11000;			
RX	MEDLINE-21681879; PubMed-11823852;			
RA	Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,			
RA	Arlat M., Billault A., Brotier P., Camus J.C., Catolico L.,			
RA	Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,			
RA	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,			
RA	Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,			
RA	Weissenbach J., Boucher C.A.,			
RT	"Genome sequence of the plant pathogen Ralstonia solanacearum.";			
RL	Nature 415:497-502(2002).			
DR	EMBL: AL646078; CAD17473.1; -			
DR	InterPro: IPR002034; AIPM/HcIt_synth.			
DR	InterPro: IPR000891; HMGL-like.			
DR	Pfam: PF00682; HMGL-like; 1.			
DR	TIGRFAMs: TIGR00970; leuA_yeast; 1.			
DR	PROSITE: PS00815; AIPM_HOMOCIT_SYNTH_1; UNKNOWN_1.			
DR	PROSITE: PS00816; AIPM_HOMOCIT_SYNTH_2; UNKNOWN_1.			
KW	Lyase; Plasmid; Complete proteome.			
SQ	SEQUENCE 576 AA; 63149 MW; BBGBDA9A66BA332B CRC64;			
Query Match	39.38;	Score 55;	DB 16;	Length 576;
Best Local Similarity	44.48;	Pred. No. 4.4;		
Matches 12;	Conservative 3;	Mismatches 12;	Indels 0;	Gaps 0;

QY 1 RLEAKHRENVGHERMGRGRTSSKELA 27  
 ||| : ||| : ||| : |||  
 DB 550 RLAKRHQAQPGQSLIRGMAPSMELA 576

## RESULT 2

Q9HTU8 PRELIMINARY; PRT; 157 AA.  
 ID 09HTU8  
 AC 09HTU8  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical protein PA5246.  
 GN PA5246.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 [1]

SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 15692 / PA01;  
 MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Madman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Kretzer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004937; AAC08631.1; -  
 DR InterPro: IPR003736; DUF157.  
 DR Pfam: PF02584; DUF157; 1.  
 DR TIGRFAMs: TIGR00369; unchar\_dom\_1; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 157 AA; 17005 MW; 1002F16B858253A3 CRC64;

Query Match 38.9%; Score 54.5; DB 16; Length 157;  
 Best Local Similarity 40.0%; Pred. No. 1.3;  
 Matches 14; Conservative 2; Mismatches 0; Indels 19; Gaps 2;

QY 5 KHRENVGHERM-----GGRGT 21  
 :||| :||| :||| :|||  
 DB 85 RHRE-LPHERMARLSKLTIDRLVYLRPGRGRT 118

## BUILT 3

5SP6 PRELIMINARY; PRT; 251 AA.  
 ID 095SF6  
 AC 095SF6  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE GH28833p.  
 GN CG2640.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Gelinker S.,  
 RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AY060818; ALU28366.1; -  
 DR FlyBase; FBgn0037474; CG2640.

SQ SEQUENCE 251 AA; 28630 MW; C7CF5BD49E079P9 CRC64;

Query Match 36.1%; Score 50.5; DB 5; Length 251;  
 Best Local Similarity 48.0%; Pred. No. 9;  
 Matches 12; Conservative 3; Mismatches 3; Indels 7; Gaps 1;

QY 10 VPGHERMGRGRTSS-----KELA 27  
 :||| :||| :||| :|||

DB 153 LPHERGLGRSLGSLQTELEKELA 177

## RESULT 4

Q9YI66 PRELIMINARY; PRT; 491 AA.  
 ID 09YI66  
 AC 09YI66  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE CG2640 protein.  
 GN CG2640.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Gelinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitai M., Kaush G., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheelers F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003672; AAF54059.1; -  
 DR FlyBase; FBgn0037474; CG2640.  
 SQ SEQUENCE 491 AA; 55811 MW; 59AE06D414EAAE1B CRC64;

Query Match 36.1%; Score 50.5; DB 5; Length 491;  
 Best Local Similarity 48.0%; Pred. No. 19;

Matches 12; Conservative 3; Mismatches 3; Indels 7; Gaps 1;  
QY 10 VPGHERMGRGRTSS-----KELA 27  
: ||||: |||: ||||  
Db 171 LPsherLGRSLSQFTIELTKELA 195

RESULT 5  
Q9CIP3 PRELIMINARY; PRT; 72 AA.  
AC Q9CIP3; 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
GN Hypothetical protein ydbC.  
OS YABC OR L10313.  
OC Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Lactococcus.  
NCBI\_TaxID=1360;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-111403;  
RA MEDLINE-2135186; PubMed-11337471;  
RA Bolotin A., Winkler P., Manger S., Jalllon O., Malarne K.,  
RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
RT lactis ssp. lactis 111403";  
RL Genome Res. 11:731-753(2001).  
DR EMBL: AB006268; AAK0411.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 72 AA; 8405 MW; C5041FC4543B8410 CRC64;

Query Match 35.7%; Score 50; DB 16; Length 72;  
Best Local Similarity 60.0%; Pred. No. 2.8;  
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 11 PGHERMGRGRTSKE 25  
: ||||: |||: ||||  
Db 46 PDHEKMGKGTLSRE 60

RESULT 6  
Q9XIX9 PRELIMINARY; PRT; 838 AA.  
AC Q9XIX9; 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)  
DE 01-NOV-1999 (TREMBlrel. 12, last annotation update)  
OS Similar to Herpesvirus papio BRRF2 homolog gene.  
OC Oryza sativa (rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
NCBI\_TaxID=4530;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-CV. NIPPONBARE;  
RA Sasaki T., Nagamura Y., Yamamoto K.;  
RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 6, PAC  
RT clone: P0681P10.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB026295; BAA81852.1; -; 5E14371CB58B461F CRC64;  
SQ SEQUENCE 838 AA; 92789 MW; 5E14371CB58B461F CRC64;

Query Match 35.7%; Score 50; DB 10; Length 838;  
Best Local Similarity 48.0%; Pred. No. 40;  
Matches 12; Conservative 1; Mismatches 12; Indels 0; Gaps 0;  
QY 1 RLEAKHENVPGHERMGRGRTSKE 25  
: ||||: |||: ||||  
Db 191 RTRAKGERLTLGRGRGRGLSME 215

RESULT 7  
Q74328 PRELIMINARY; PRT; 424 AA.  
AC Q74328; Q9UGS2;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DE 01-NOV-1998 (TREMBlrel. 08, last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
DE Hypothetical 47.0 kDa protein Cl685.08 in chromosome II.  
GN SPCL685.08.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
NCBI\_TaxID=4896;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA Wood V., Rajandream M.A., Barrell B.G., Gilbert H., Duesterhoeft A.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

Query Match 35.4%; Score 49.5; DB 3; Length 424;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 12; Conservative 3; Mismatches 6; Indels 3; Gaps 1;  
QY 4 AKHENVPGHERMGRGRTSKELA 27  
: ||||: |||: ||||  
Db 169 AKESLIP---RRSGRTSSKSLs 189

RESULT 8  
Q002316 PRELIMINARY; PRT; 2140 AA.  
AC Q002316; 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 09, last sequence update)  
DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
DE Mature fruiting body-specific cell-adhesion protein (MFBA).  
GN MFBA OR MFBI.  
OS Lentinula edodes (Shiitake mushroom) (Lentinus edodes).  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
OC Agaricales; Tricholomataceae; Lentinula.  
NCBI\_TaxID=5353;  
[1]  
SEQUENCE FROM N.A., AND SEQUENCE OF 565-570.  
RC STRAIN-FMC2;  
RA MEDLINE-95172398; PubMed-7867945;  
RA Konoh O., Muto A., Kajiwara S., Takagi J., Saito Y., Shishido K.;  
RT "A fruiting body-specific cDNA, mfbA, from the mushroom Lentinus  
RT edodes encodes a high-molecular-weight cell-adhesion protein  
RT containing an Arg-Gly-Asp motif.";  
RL Gene 154:31-37(1995).  
CC - FUNCTION: MAY PLAY A ROLE IN CELL ADHESION.  
CC - TISSUE SPECIFICITY: FRUITING BODY-SPECIFIC. HIGHLY EXPRESSED IN  
CC THE PILEUS, MODERATE EXPRESSION IN THE STIPE, LOW LEVELS OF  
CC EXPRESSION IN THE GILLS.  
CC - DEVELOPMENTAL STAGE: EXPRESSION IS HIGH IN MATURE FRUITING BODIES,  
CC DETECTABLE IN IMMATURE FRUITING BODIES BUT ABSENT IN EARLIER

CC STAGES OF DEVELOPMENT, AND IN THE VEGETATIVE MYCELIUM.  
 CC -1- STABILITY: CONTAINS 1 'SET' DOMAIN.  
 DR EMBL: D14487: BA18903.1; -  
 DR InterPro: IPR001214; SET.  
 DR InterPro: IPR001965; Znf\_PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00249; PHD; 1.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PSS0280; SET; 1.  
 DR Cell adhesion.  
 KW CYS-RICH (PHD-FINGER).  
 FT DOMAIN 144 195  
 FT SET DOMAIN.  
 FT SITE 376 548  
 FT CELL ATTACHMENT SITE (POTENTIAL).  
 SQ SEQUENCE 2140 AA; 232670 MW; 95CB36DE3A77C8FB CRC64;  
 Query Match 35.4%; Score 49.5; DB 3; Length 2140;  
 Best Local Similarity 33.3%; Pred. No. 1.3e+02;  
 Matches 12; Conservative 6; Mismatches 7; Indels 11; Gaps 1;  
 3 EAKHRENVPGH-----ERMGRGRTSSKELA 27  
 |||::|||:|  
 Db 159 EAKDENLDPHMCWECPSLIERIDREKANSLOTA 194  
 RESULT 9  
 Q9V712 PRELIMINARY: PRT; 193 AA.  
 AC Q9V712;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE CG13941 protein.  
 GN CG13941.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCB1\_Taxid=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.E., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de la Piedad G., Deleter A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Maltsev B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003815; AAF58254.1; -  
 DR FlyBase: FBgn0033928; CG13941.  
 DR InterPro: IPR005162; Retrotrans\_gag.  
 DR Pfam: PF03732; Retrotrans\_gag; 1.  
 SQ SEQUENCE 193 AA; 22621 MW; 5841D186C065D391 CRC64;  
 Query Match 35.0%; Score 49; DB 5; Length 193;  
 Best Local Similarity 38.5%; Pred. No. 12;  
 Matches 10; Conservative 6; Mismatches 4; Indels 6; Gaps 1;  
 QY 2 LEAKHRENVPGH-----ERMGRGRT 21  
 ::|||::|  
 Db 162 MQPKYRESIPREHVKTFRELDGRGT 187  
 RESULT 10  
 Q8YOK8 PRELIMINARY: PRT; 829 AA.  
 AC Q8YOK8;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein A1R3815.  
 GN A1R3815.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCB1\_Taxid=103690;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kurihara T., Sasamoto S.,  
 Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 Nakazaki N., Shimizu S., Sugimoto M., Takazawa M., Yamada M.,  
 Yasuda M., Tabata S.;  
 RA "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120."  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL: AP003594; BAB75514.1; -  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR001300; Protease\_C2.  
 DR Pfam: PF00648; Peptidase\_C2; 1.  
 DR SMART: SM00230; Cyspc; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN.1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 829 AA; 89203 MW; 8B2D095E2731AE33 CRC64;  
 Query Match 35.0%; Score 49; DB 16; Length 829;  
 Best Local Similarity 45.0%; Pred. No. 56;  
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 QY 7 RENVGHERMGRGRTSSKEL 26  
 |||::|  
 Db 704 RENTPGYNLIGNDPLNRL 723  
 RESULT 11  
 Q8SB46 PRELIMINARY: PRT; 483 AA.  
 AC Q8SB46;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative polyprotein.  
 GN OSJNB0091009.13.



OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Euphorbiaceae; Oryzae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
 RT Sasaki C., Henry D., Oates R., Simmons J.;  
 "Rice Genomic Sequence."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC091732; AAL7162.1; -  
 KW Polyprotein.  
 SQ SEQUENCE 483 AA; 55242 MW; D2B0A0BCB6CFADB CRC64;

Query Match 34.6%; Score 48.5; DB 10; Length 483;  
 Best Local Similarity 45.8%; Pred. No. 38;  
 Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

4 AKHRENVPGH-ERMGRGRTSSKEL 26  
 107 AKHNAVYHPQTSGRATSNKQI 130

## RESULT 12

09K5L1 PRELIMINARY; PRT; 765 AA.

AC 09K5L1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
 DE Comp.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RS-B-1;  
 RA Tortosa P., Logsdon L., Kraigher B., Itoh Y., Mandic-Mulec I.,  
 RA Dubnau D.;  
 "Specificity and genetic polymorphism of the Bacillus competence  
 quorum-sensing system."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY003900; AAF82174.1; -  
 DR InterPro: IPR003594; ATPbind\_ATPase.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR001478; PDZ.  
 PFam: PF00595; PDZ.1.  
 SMART: SM00387; HATPase\_c; 1.  
 DR SMART; SM00228; PDZ; 1.  
 DR PROSITE; PS50106; PDZ; 1.  
 SQ SEQUENCE 765 AA; 88476 MW; 8118331678A6AC83 CRC64;

Query Match 34.6%; Score 48.5; DB 2; Length 765;  
 Best Local Similarity 44.4%; Pred. No. 62;  
 Matches 12; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

QY 1 RLEAKHRENVPGHERMGRGRTSSKEL 26  
 645 KLVAAQOQERVPFHRLNTRFTASIDL 671

## RESULT 13

09K5K7 PRELIMINARY; PRT; 765 AA.

AC 09K5K7;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
 DE Comp.  
 OS Bacillus mojavensis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=72360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tortosa P., Logsdon L., Kraigher B., Itoh Y., Mandic-Mulec I.,  
 RA Dubnau D.;  
 "Specificity and genetic polymorphism of the Bacillus competence  
 quorum-sensing system."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY003901; AAF82176.1; -  
 DR InterPro: IPR003594; ATPbind\_ATPase.  
 DR InterPro: IPR003661; HIS\_KIN.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR001478; PDZ.  
 PFam: PF00595; PDZ.1.  
 SMART: SM00387; HATPase\_c; 1.  
 DR SMART; SM00388; HSKA; 1.  
 DR SMART; SM00228; PDZ; 1.  
 DR PROSITE; PS50106; PDZ; 1.  
 SQ SEQUENCE 765 AA; 88605 MW; 725552C6A8B76A62 CRC64;

Query Match 34.6%; Score 48.5; DB 2; Length 765;  
 Best Local Similarity 44.4%; Pred. No. 62;  
 Matches 12; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

QY 1 RLEAKHRENVPGHERMGRGRTSSKEL 26  
 645 KLVAAQOQERVPFHRLNTRFTASIDL 671

## RESULT 14

08TF63 PRELIMINARY; PRT; 244 AA.

AC 08TF63;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)  
 DE DCPN1.  
 GN DCPN1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21656978; PubMed=11798177;  
 RA Masuda M., Senju S., Fujii S., Terasaki Y., Takeya M., Hashimoto S.,  
 RA Matsushima K., Yumoto E., Nishimura Y.;  
 "Identification and Immunocytochemical Analysis of DCPN1, a Dendritic  
 Cell-Associated Nuclear Protein."  
 RL Biochem. Biophys. Res. Commun. 290:1022-1029(2002).  
 DR EMBL; AB074498; BAB84585.1; -  
 SQ SEQUENCE 244 AA; 26704 MW; 803125CEFA61B9F8 CRC64;

Query Match 34.3%; Score 48; DB 4; Length 244;  
 Best Local Similarity 66.7%; Pred. No. 21;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 ENVPGERMGRG 19  
 18 ETVPGHORLERG 29

## RESULT 15

094827 PRELIMINARY; PRT; 864 AA.

AC 094827;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)  
 DE KIA0720 protein (Fragment).

GN KIAA0720.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=99087487; PubMed=9872452;  
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RT \*Prediction of the coding sequences of unidentified human genes. XI.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.\*;  
 RL DNA Res. 5:277-286(1998).  
 DR EMBL; AB018263; BAA34440.1; -;  
 DR InterPro; IPR001849; PH. RhGEF.  
 DR InterPro; IPR000219; RhGEF.  
 DR Pfam; PF00169; PH.1.  
 DR Pfam; PF00821; RhGEF; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00325; RhGEF; 1.  
 DR PROSITE; PSS0003; PH\_DOMAIN; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 864 AA; 95282 MW; 1EA2E7DB43E595D7 CRC64;  
 Query Match 34.3%; Score 48; DB 4; Length 864;  
 Best Local Similarity 40.0%; Pred. No. 84;  
 Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 RLEAKHRENVGHERMGRTSKE 25  
 ||| ||: |||: | : |  
 Db 173 RLEDSWRELIDGHEKLTTRQCHQE 197

Search completed: October 29, 2002, 13:57:35  
 Job time : 80 secs